nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



VALIUM'S TARGET

Structure of $GABA_A$ receptor reveals binding sites for anti-anxiety drugs PAGES 37 & 67

EVALUATION

SCIENCE FOR SOCIETY

A fresh way to judge work aimed at changing lives

PAGE 23

CELL BIOLOGY

DNA REPAIR KIT

Polymerized actin helps to maintain genome integrity

PAGES 35, 54 & 61

ASTRONOMY

TESTING EINSTEIN

Triple-star system adds weight to theory of gravity

PAGES 40 & 73

NATURE.COM/NATURE

5 July 2018

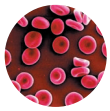
Vol. 559, No. 7712

THIS WEEK

EDITORIALS

WELCOME New *Nature* editor Magdalena Skipper offers a glimpse of the future **p.6**

WORLD VIEW Harness the secrets of ecosystems to tackle urban problems **p.7**



FLEXIBLE FRIENDS Red blood cells offer a new route to making plastics **p.9**

Overdose response

Information about the structure of a key brain protein could help in the development of safer prescription drugs.

The musician Tom Petty last year joined a long line of high-profile stars to hit the headlines for dying from an accidental overdose of prescription painkillers and tranquillizers. The risks of opioid painkillers dominate the news. But the other half of the deadly mixture in such overdose cases deserves attention, too. It's typically a medicine from a class of drugs called benzodiazepines or — more colloquially — benzos.

These drugs — Valium (diazepam), Xanax (alprazolam) and the like — have been prescribed as tranquillizers for decades, mainly for anxiety. And the health risks they pose, such as addiction, are well known. Yet a study published last week spells out the specific danger of mixing them with an opioid painkiller: the risk of an overdose is five times higher when people taking opioids are also prescribed a benzodiazepine (I. Hernandez *et al. JAMA Network Open* **1**, e180919; 2018). According to the US National Institutes of Health's National Institute of Drug Abuse, almost one-third of overdoses involving opioids also involve benzos.

The problems come because both opioids and benzos function as depressants on the central nervous system. Acting together, they can make it difficult to breathe — so difficult, in fact, that many overdose deaths are caused by suffocation. As a result, the US Food and Drug Administration (FDA), among other regulatory bodies, has repeatedly tried to highlight the risks of prescriptions for both types of drug. In a perspective article published last month, the FDA called for "proactive pharmacovigilance" — better tracking of the drugs that different (and sometimes the same) doctors prescribe to patients (D. C. Throckmorton *et al. N. Engl. J. Med.* http://doi.org/gdj5t4; 2018).

A better long-term solution would be safer drugs. Work is under way already to develop non-addictive opioids. Also on the radar is replacements for benzos, with the aim of finding compounds with the same calming effect but none of the risky side effects.

In *Nature* this week scientists present an advance that could help in this search: the detailed structure of a receptor known as $GABA_A$, a membrane protein that is found mainly near the brain's synapses (see page 67). The physiological effects of benzos and some other drugs are triggered by binding to this receptor. Knowing its structure might help researchers to design or identify less-risky compounds. Alternative drugs are a long way down the road, but chemists are already finding compounds that can bind to other receptors with known structures. Information about exactly how benzo drugs bind to this receptor should provide insights to help in the development of new compounds for the treatment of anxiety, epilepsy and other neurological disorders.

The structure of the $GABA_A$ receptor is so complicated that chemists had been unable to solve it using conventional structural-biology techniques. Instead, they turned to cryo-electron microscopy (cryo-EM), a technique that fires beams of electrons at proteins that have been frozen in solution, and that is rapidly overtaking conventional techniques for solving the structures of the more complicated human proteins.

The $GABA_A$ receptor comprises several subunits, which can be arranged in a number of ways in what are called isoforms; the *Nature* paper looks at the main isoform found at human brain synapses. The high-resolution structures reveal how the neurotransmitter GABA (γ-aminobutyric acid) and benzodiazepines such as flumazenil, bind to interfaces between the subunits.

"The complexity of the protein could help to explain one of its most striking features: its versatility."

The complexity of the protein could help to explain one of its most striking features: its versatility. As well as binding to benzos, the receptor is also activated by compounds such as barbiturates, alcohol, anaesthetics and steroids. So knowing the structure should help chemists to find therapeutic compounds in those fields, too. The next steps will probably involve working out the modes of action of these chemically distinct drug classes.

The original medicines have helped millions of people and, used with care, can continue to do so. But too much of a good thing can and does kill people. And that needs to stop. Fundamental structural biology — once seen as a purely theoretical pursuit — can point the way. ■

City living

Researchers must get to grips with the grave challenge of rapidly growing urbanization.

Cape Town has avoided its Day Zero, when it was due to run out of water, but has instead ended up with scenes reminiscent of the disaster film *The Day after Tomorrow*. Heavy winter rains in the South African city this week caused widespread flooding. Dry for so long, the city's storm drains have been overwhelmed, and many roads have turned to rivers. Its citizens have been forced to shift from fears of a lack of water to dealing with the chaos caused by having too much.

In some ways, the experience of the city and its residents is a parable for our times, and an example of the kinds of problems faced by city officials the world over. How can we enable our urban centres to cope with extremes of weather, including those that global warming is expected to bring? And how can planning for worst-case scenarios be aligned with increasing awareness of the need to make our cities more sustainable?

That need is great. It's well documented that humans are becoming an urban and not a rural species, and that trend is expected to continue. In 1950, less than one-third of the population lived in cities. More than 50% do so now, and by 2050 the figure is projected to be 70%. Every week, the urban population grows by around 1.5 million. By

© 2018 Springer Nature Limited. All rights reserved.

mid-century, as many people will be living in cities as occupy the entire planet today. About one-third of this new urban population growth will be in just three countries — Nigeria, China and India.

All of those billions of city-dwellers have entirely predictable demands: food, water, and subsequent waste disposal. Realistically, most world cities will struggle to become self-sufficient when it comes to dealing with those basic needs of their citizens. High-density housing and infrastructure don't leave much space for fields and landfill. So the idea of a truly sustainable city is something of a misnomer. But to make cities more sustainable than they are now is a noble goal, and one included in the United Nations' Sustainable Development Goals (SDGs) agreed in 2015.

Specifically, SDG 11 pledges to "make cities and human settlements inclusive, safe, resilient and sustainable". Progress towards that goal will be discussed at a meeting — the High-Level Political Forum on Sustainable Development — at UN headquarters in New York City later this month. Experts in the field will also meet in Singapore next week for a one-day Springer Nature event on Science and the Sustainable City (go.nature.com/2mluyjd).

In a World View article this week (see page 7), one of the conference speakers makes the case for city leaders to take an 'ecosystem' approach to urban environmental management. Most cities have exploited the surrounding ecosystems to support their needs, but efforts to make this relationship less damaging — and so more sustainable — are crucial. Cities are ecosystems, too, of course, and the explosion of urban science to model and understand these systems — another topic to be discussed at the Singapore meeting — demonstrates one important way in which research can contribute.

It's clear that approaches, including research, must be tailored to the individual needs of cities and the regions around them, wherever they are. But, in every case, rampant urbanization demands integrated and multidisciplinary thinking.

Delegates to the New York meeting need only look around them to see one emerging problem with SDG 11 that demonstrates the point. Experts are questioning whether "resilient" and "sustainable" are as complementary as they at first seem, or whether they might instead conflict. To help make the Big Apple more resilient to rising sea levels, plans for a massive, retractable storm-surge barrier across New York Harbor were floated after the devastation caused by Hurricane Sandy in 2012. But others have pointed out that a US$20-billion project that could damage ecosystems, and that would almost never be used, can hardly be classed as sustainable. The same debate is unfolding in Cape Town. Although many blame climate change for the drought there, experts argue that the root cause of the problem is more straightforward: reluctance among policymakers to invest in a new reservoir system. In part, this failure to make the city more resilient was caused by concerns that the dams required were not a sustainable way to do it.

"Rampant urbanization demands integrated and multidisciplinary thinking."

Such decisions and choices are difficult because they require weighing up often conflicting demands and priorities. Science can help by gathering and analysing data to define risks and benefits. Researchers can thus help to tackle these thorny problems, and to ensure that cities continue to offer a home to people — and to much of science itself. ■

Opening remarks

Nature's eighth editor-in-chief extends a welcome to readers and outlines her vision.

Last week, Philip Campbell stepped down as the editor-in-chief of *Nature* after 22 years, and he passed the baton to me.

It is with great honour and sense of responsibility that I take on this unique role, as only the eighth editor-in-chief in *Nature*'s history. I am excited by the prospect of serving the scientific community from this new perspective, bringing to it my own varied editorial experience.

Phil began his parting words in last week's issue by restating that *Nature*'s editorial role has been consistently about support for outstanding science. True to our mission, we continue to enable dissemination of science. It has always been at the very heart of everything we do to facilitate communication of outstanding discoveries that are important and relevant to society and the world, as well as to the scientific community itself.

We will continue to widen our horizons at *Nature*, to be the venue for not only the established, but also emerging fields and discoveries that defy traditional discipline boundaries. There will always be room for the publication and coverage of fundamental discoveries and cutting-edge applications alike. And through our primary research papers, news and analyses, we will continue to further the societal impact of science.

Tomorrow's discoveries will be made by today's early-career researchers. I believe we at *Nature* have an important part to play in their mentorship. We hope to engage them more fully in the dissemination and vetting of the scientific record — for example, by promoting peer-review mentorship. We also hope to be guided by them, so that we can meet their needs when it comes to publishing their work, as research becomes more data-rich and computationally heavy.

In the past, *Nature* has adapted to accommodate the evolving nature of discoveries. In future, I will be keen to explore ways in which our formats can adapt further as research requirements continue to evolve.

The journal will continue to work with research communities to enhance reproducibility and transparency in science. As researchers aim for more transparency in how they both design studies and generate and analyse data, I hope we will make our own processes increasingly transparent, with the aim of demystifying the way we work.

Much has already been made of my being the first female editor-in-chief of *Nature*. Although this may well have been overdue, it is a reflection of greater diversity in the research community, and not just in terms of gender. With human diversity comes diversity of priorities, views and interests, all of which deserve equal prominence. We will continue to strive towards equal representation of all groups among our authors, reviewers and staff. Diversity of collaborators will enhance diversity in our coverage of topics. We know we have some way to go — for example, to increase the number of female reviewers — and we will continue to make it our priority.

The fact that I am one of only two life scientists to hold the title of editor-in-chief of this journal has attracted less attention. My background undoubtedly means that I bring a particular perspective on the sciences, although my focus will remain broad. To my eyes as a geneticist, the most influential work published in *Nature*'s pages under the watch of my predecessor — a physicist — must surely be the reporting of the human genome sequence. Under my watch, I look forward to publishing equally influential findings quite removed from my own area of research expertise.

The role of editor-in-chief is not unlike that of an orchestra's conductor. Behind every piece of journalism and every paper there are dedicated and talented professionals; experts who know how to play their part. I very much look forward to enhancing this well-honed performance, not just at *Nature* magazine itself but also in the broader family of Nature Research journals.

Nature is your journal. Without the scientific community, we have no authors, no reviewers and no readers. I look forward to learning from you all, so that I and *Nature* may serve you better. **Magdalena Skipper**

© 2018 Springer Nature Limited. All rights reserved.



CLIVE HILLIKER

Advance the ecosystem approach in cities

To make cities better for people and the planet, the approach must become more systemic and pragmatic, and include the global south, says **Xuemei Bai**.

More than half of the world's population now live in urban areas, and both the proportion and numbers continue to grow. In future, we will require drastically different ways of planning, building and governing cities.

Since at least the 1970s, researchers have been doing long-term studies to understand how cities alter ecosystems. Now, the lens is on what ecology can do for cities. The ecosystem approach incorporates nature into urban settings to make them more sustainable, liveable and resilient, and means managing cities themselves as ecosystems: intricately connected, dynamic subsystems of social, built and natural components. The approach encompasses concepts such as green infrastructure, nature-based solutions (adopting sustainable practices that harness the natural world) and ecosystem-based adaptation (managing natural ecosystems to help cities face climate change).

One of many compelling demonstration projects includes the Vertical Forest (*Bosco Verticale*) in Milan, Italy. There, more than 900 trees and thousands of shrubs and other flora have been planted on the facades of two residential towers to improve air quality and carbon storage. The benefits have been estimated as equalling those offered by 2 hectares of natural forest. Meanwhile, 'sponge cities' are being built across China to mitigate increasingly frequent and severe flooding. Roofs are being covered with vegetation, wetlands restored and roads and other spaces designed with permeable surfaces. Urban managers hope that these will absorb enough rainwater to prevent major floods.

Sundry individual projects reveal the benefits of the ecosystem approach, including water purification, improved public health, reduced disaster exposure, enhanced resilience and social justice, but it will take a lot more to change the future of cities and achieve the global targets agreed to in the United Nations' Sustainable Development Goals and emerging guidelines encompassed in the New Urban Agenda.

Several hurdles must be cleared: researchers and practitioners must dissect what hasn't worked, develop approaches suited to the global south and learn to unite disparate efforts in a systemic way.

To start with, studies need to go beyond individual success stories. Researchers must ask when, and to what extent, the ecosystem approach can truly replace conventional infrastructure. If it is to be incorporated into urban structure seriously, its effectiveness, constraints and scalability must be discussed openly and compared with mainstream engineering solutions. For example, green rooftops can require considerable water resources; where water is scarce, it is essential to invest in technology to reuse 'grey water', or waste water from households and offices. The fact that such critiques are hard to find in the literature signals how the ecosystem approach is still at an early stage of application in practice.

In addition, the ecosystem approach must be assessed and applied in the global south. The vast majority of studies are in cities in Europe or North America, with a few from China. But the greatest challenges lie elsewhere in Africa, where urbanization is proceeding particularly fast, and in Asia's and Latin America's many developing countries.

Here, priorities and appropriate solutions will differ. For example, Amsterdam invested €20 million (US$23 million) — and leveraged up to €55 million in co-financing — for its 2015–18 Green Agenda. Developing cities, with their limited funds, often need first to meet more-severe and urgent infrastructure needs, such as sanitation and water supply.

I have seen this at work. Just after finishing my PhD at the University of Tokyo, I took on a reforestation project at a mining site in Ma'anshan, China. The approach had proved successful at about 1,400 experimental sites in Japan. However, our local counterparts saw no point in planting native trees that they perceived as neither beautiful nor useful. They were also concerned about the cost of maintaining and monitoring the reforested area. Eventually, we developed a plan to ring the site with fruit trees. These supplied revenue that supported workers and enabled the project's success. Both researchers and practitioners must learn the art of adaptation.

THE ECOSYSTEM APPROACH INCORPORATES NATURE INTO URBAN SETTINGS, AND MANAGES CITIES AS ECOSYSTEMS.

Most of all, integration is essential. Right now, projects are labelled as 'green' for adding vegetation, 'blue' for bodies of water and 'grey' for built-up systems. We need more 'kaleidoscopes' that amalgamate all of these elements and more. Cities must be viewed as a human-dominant ecosystem, with natural, social, technical, economic and cultural components, and we need to apply systems principles — a recognized approach in urban ecology — to manage them.

Analysis and planning must bring together different city functions, weigh trade-offs and hunt out synergies. We researchers need to get our acts together by bridging conceptual, methodological and disciplinary divides to develop integrated and actionable urban knowledge. Sometimes, the same stretch of road is dug up multiple times, by different city departments working independently to improve the water, sewage and gas pipes and road quality. City planners often try to optimize just one variable without considering others, for example, designing a suburb to fit in a lot of residents, without providing much room for green spaces. High density might be the right direction, but it needs to be examined alongside other implications and trade-offs.

To move cities — and so the planet — into a sustainable future, we need to break down silos and learn how to improvise in context. ■ SEE EDITORIAL P.5

Xuemei Bai *is a professor at the Fenner School of Environment and Society at the Australian National University in Canberra.*
e-mail: xuemei.bai@anu.edu.au

© 2018 Springer Nature Limited. All rights reserved.

SEVEN DAYS *The news in brief*

FACILITIES

Equipment damaged

Two of the world's most-powerful electron microscopes sustained significant damage when a magnitude-6.1 earthquake hit the Japanese coastal city of Osaka on 18 June. The event killed 3 people and injured more than 200. Hidehiro Yasuda, director of Osaka University's Research Center for Ultra-High Voltage Electron Microscopy, told Japanese media last week that the quake damaged the circuits and high-voltage tank of the centre's 12-metre-high H3000 ultra-high-voltage electron microscope, one of the world's largest electron microscopes. Also damaged was the centre's materials- and bio-science electron microscope, which is used to study biological material. Yasuda said that research programmes that use the equipment could be affected for at least a year.

ENVIRONMENT

Lost trees

Tree cover equivalent to nearly 16 million hectares of forest disappeared across the tropics in 2017, marking the second-worst year for such loss on record, according to a 26 June announcement from Global Forest Watch, an environmental group based in Washington DC. The organization used new satellite data from the University of Maryland in College Park for their analysis. Likely causes for the destruction include forest fires in the Brazilian Amazon and an increase in deforestation in Colombia. Hurricanes wiped out trees in the Caribbean, and agriculture, charcoal production and logging drove losses in the Democratic Republic of the Congo.



IAN BOTTLE/ALAMY

Belize reef removed from danger list

The United Nations World Heritage Committee has removed the Belize Barrier Reef Reserve System from its List of World Heritage in Danger, the group announced on 26 June. The barrier reef, the largest in the Northern Hemisphere, includes atolls, coastal lagoons and mangrove forests, and is home to several threatened marine species, such as the West Indian manatee (*Trichechus manatus*). The committee added the area to the list of endangered sites in 2009 following concerns about marine habitat destruction, offshore oil extraction and other issues. The decision to delist the site came as a result of conservation efforts, including a ban on oil exploration in Belize's waters.

EVENTS

Climate lawsuit

A US federal judge threw out a high-profile lawsuit filed against fossil-fuel companies by two California cities, according to news reports on 26 June. San Francisco and Oakland were suing several companies — including BP, Chevron, Exxon Mobil and Royal Dutch Shell — in an effort to recoup some of the costs of dealing with the adverse impacts of climate change. In a decision issued on 25 June, judge William Alsup of the federal district court in San Francisco acknowledged climate-change science, but stated that the issue should be addressed by the government's executive branch and Congress, not the courts. The decision raises questions about similar lawsuits filed by other local governments across the United States.

RESEARCH

Misconduct inquiry

The Karolinska Institute in Stockholm has declared seven researchers responsible for scientific misconduct in a case involving six research articles co-authored by disgraced thoracic surgeon Paolo Macchiarini. It says that Macchiarini, who transplanted synthetic windpipes into three patients at the Karolinska University Hospital between 2011 and 2013, held ultimate responsibility for the papers. The institute's investigation found that the studies included "fabricated and distorted descriptions of the patients' conditions". The experimental procedures failed. The president of the Karolinska Institute, Ole Petter Ottersen, has called for the papers to be retracted. The outcome overturns a 2015 decision from an investigation under the previous president, which stated that Macchiarini had not committed misconduct in these papers. The institute said in a press

© 2018 Springer Nature Limited. All rights reserved.

release on 25 June that a further 31 co-authors are "blameworthy for their contributions" to the articles, but not guilty of misconduct. Another five co-authors were cleared of all blame. See go.nature.com/2kw93zu for more.

PEOPLE

Biologist resigns

An internal investigation by the University of California, Irvine (UCI), has substantiated sexual-harassment claims against evolutionary biologist Francisco J. Ayala, according to a 28 June statement from UCI chancellor Howard Gillman. Ayala, a prominent donor to UCI whose name was on the School of Biological Sciences, resigned from the university, effective 1 July. Gillman announced that UCI would remove Ayala's name from the biology department, the science library and various programmes that bear his name. In a statement, Ayala said that it was never his intent to make his colleagues uncomfortable. The claims that prompted the current investigation were not the first, says Micha Liberty, an attorney representing three of the four women involved in the matter. The university received an official complaint against Ayala three years ago, she says.



JAXA, UNIV. TOKYO AND COLLAB.

SPACE

Rocky rendezvous

Japan's Hayabusa-2 spacecraft has arrived at the asteroid Ryugu (**pictured**), kicking off a year and a half of exploration that will culminate with the return of rock samples to Earth. The Japan Aerospace Exploration Agency said on 27 June that the probe had reached its 'rendezvous' with the 1-kilometre-wide asteroid, after its chemical thrusters performed a series of automated braking burns. Hayabusa-2 is now hovering 20 kilometres above Ryugu, preparing to release four landers and to make a series of touchdowns itself. If all goes to plan, the mission will be the second — after its predecessor, Hayabusa-1 — to return an asteroid sample to Earth.

POLICY

Three-person IVF

A group of Australian politicians has released a road map for the country to move towards legalizing mitochondrial donation. The group's recommendations, published on 27 June, include that the government consult the public and scientific experts about permitting clinical use of the reproductive technology, which could help women avoid passing genetic defects to their children through mutations in cellular organelles called mitochondria. The technique uses a healthy donor egg to create an embryo with the nuclear DNA of two people and the mitochondrial DNA of the healthy egg. The embryo can then be implanted using *in vitro* fertilization (IVF). The United Kingdom is currently the only country to specifically allow the technology. The group of conditions called mitochondrial disease affects as many as one in 500 children, and symptoms range from mild to severely debilitating and life-threatening, affecting many systems in the body. See go.nature.com/2kligdy for more.

Weapons watch

The world's chemical-weapons watchdog now has the authority to attribute responsibility for chemical attacks. The Organisation for the Prohibition of Chemical Weapons (OPCW) implements the Chemical Weapons Convention, which bans the production and use of such arms. Until now, the body could provide independent laboratory confirmation of the chemical used in an attack, but could not attribute a likely origin. After a resurgence in chemical-weapons attacks in Syria in the past few years, however, the OPCW has increasingly taken on investigative tasks, and, on 27 June, most of the convention's member states voted in favour of expanding its power. To carry out its new role, the OPCW will need to increase its infrastructure and forensics capacity, which will help it to build up 'signatures' of chemical components and provide investigators with clues about where the chemicals were made — and by whom. The organization, based in The Hague, the Netherlands, also oversees and verifies the destruction of existing weapons stockpiles worldwide, and monitors the chemical industry to prevent the production of new ones.

Red-wolf plans

US officials have announced plans to allow the killing of up to two dozen endangered red wolves (*Canis rufus*). On 27 June, the US Fish and Wildlife Service proposed draft regulations that would limit the animals' habitat to federal land in two North Carolina counties and lift restrictions on killing any wolves that strayed from that territory. About 40 red wolves are left in the wild — all in eastern North Carolina — with roughly 200 more in captive breeding facilities around the United States. The proposal is open for public comment until 30 July, and the agency will hold a public hearing on the plan in Manteo, North Carolina, on 10 July. The regulations, if approved, would come into effect in November.

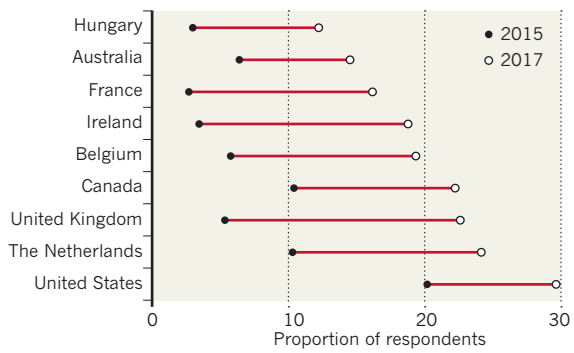
TREND WATCH

Pharmacological cognitive enhancement — the non-medical use of stimulant drugs to increase concentration or memory during work or while studying — is rising around the world, a survey of more than 100,000 people suggests. Between 2015 and 2017, the number of people who reported using illegal drugs or prescription medications for this purpose increased in all 15 nations surveyed. The most commonly used prescription drugs include modafinil, Adderall and Ritalin (methylphenidate).

SOURCE: L. J. MAIER *ET AL. INT. J. DRUG POLICY* **58**, 104–112 (2018)

QUEST FOR COGNITIVE ENHANCEMENT

Use of prescription and illegal stimulants for 'pharmacological cognitive enhancement' rose in all countries monitored by the Global Drug Survey, with European nations seeing the largest increases.



The nine countries with highest reported use in the 2017 survey.

NATURE.COM
For daily news updates see:
www.nature.com/news

© 2018 Springer Nature Limited. All rights reserved.

NEWS IN FOCUS

AGEING Study suggesting no limit on human longevity revives lifespan debate **p.14**

CYPRUS Regional research hub for climate change takes shape **p.15**

POLICY EU gets tough on export of ethically dubious research **p.17**



EMBRYOLOGY The earliest days of human development in a dish **p.19**

LIU JUNXI/XINHUA/ZUMA WIRE



About 33,000 babies have been recruited for the Born in Guangzhou Cohort Study since 2012.

EPIDEMIOLOGY

Gigantic Chinese-baby study yields rich results

Public-health insights have already emerged, and microbiome research is under way.

BY DAVID CYRANOSKI

An ambitious Chinese study tracking tens of thousands of babies and their mothers has begun to bear fruit — just six years after the study's leaders recruited their first sets of mothers and babies.

Researchers have already published results based on the cohort study, some with important public-health implications. And many more investigations are under way. One will examine infants' microbiomes, the collections of bacteria and other microorganisms that inhabit their bodies — a hot topic in health research and a key goal of the study.

The Born in Guangzhou Cohort Study[1] has recruited about 33,000 babies and their mothers since 2012. The study's leaders are hoping to reach 50,000 baby–mother sets by 2020. And this year, investigators started recruiting 5,000 maternal grandmothers to the project, enabling research across multiple generations.

"The data is vast, and there is space for many different groups globally to mine this information," says Maria Gloria Dominguez-Bello, a microbiologist at Rutgers, The State University of New Jersey, in New Brunswick, who is not involved in the study. "I really admire this effort from the Chinese team."

Ezra Susser, an epidemiologist from Columbia University in New York City, says the cohort is also important because it is tracking mothers and babies during a period of rapid economic development and social change in China, where previous studies of this type have been limited in scale.

The Guangzhou project aims to set itself apart from previous large birth-cohort studies in Norway and Denmark by enabling detailed investigations of the links between the microbiome and disease. Two others, in the United States and United Kingdom, had planned ▶

© 2018 Springer Nature Limited. All rights reserved.

to include microbiome data, but both were cancelled because of trouble recruiting participants. The US study also struggled with excessive costs and management issues.

The Chinese team has so far avoided similar problems. Its rich collection of 1.6 million biological samples includes stools, blood, placental tissue and umbilical cords. Extensive surveys also record participants' eating habits, mental health, and other lifestyle factors, such as the amount of mould in their houses.

FIRST FINDINGS

Incense burning is common in southern China, and one study based on the Guangzhou project found that exposure to the resulting fumes increases the risk of hypertension in expectant mothers[2].

Another study found that progesterone, a drug used around the world to reduce the risk of a preterm birth, was prescribed too early in pregnancy in more than 40% of women studied[3]. The researchers found that giving women the drug before 14 weeks of gestation did not reduce their chances of a preterm birth, but put them at higher risk of needing a caesarean section and of developing postpartum depression. The authors consider the findings "an urgent public-health concern".

Other studies are in progress. A team from the University of Birmingham, UK, and BGI, one of China's largest genome-sequencing institutes, in Shenzhen, is trying to characterize how the microbiomes of babies born vaginally — who are exposed to their mothers' microbes on their journey down the birth canal — differ from those of infants born by caesarean section. Although similar studies have been done on a smaller scale, Dominguez-Bello says that the Guangzhou cohort will offer statistical power to separate out other variables that could influence an infant's microbiome. These include pre- and postnatal medications and environmental pollutants.

"The data is vast and there is space for many different groups globally to mine this information."

Xiu Qiu, an epidemiologist at Guangzhou Women and Children's Medical Center and the director of the Guangzhou project, is using the cohort data to test her surprising, but tentative, finding that older mothers having a second child have a lower risk of depression during pregnancy than do women pregnant with their first child[3]. She had expected that women who already have a baby when they are pregnant would be under more stress and face a higher financial burden, and so would be more likely to experience depression. The end of China's one-child policy in 2016 means the birth-cohort study offers a fresh opportunity to study an increasing number of women, many of them older, who are having a second child, she says.

Sing Sing Way, a paediatrician at the Cincinnati Children's Hospital in Ohio, meanwhile, will be looking at the data provided by the addition of grandmothers to the study to understand why cells from mothers can live on indefinitely in their offspring. Studies in mice suggest that these cells have a protective role when the offspring are pregnant, says Way[4].

Xia Huimin, a co-founder of the project, says that the Guangzhou cohort has the power to answer many more questions like this. He hopes scientists around the world will use it. "We would like scientists from everywhere to work with us." ■

1. Qiu, X. *et al. Eur. J. Epidemiol.* **32**, 337–346 (2017).
2. He, J.-R. *et al. Sci. Total Environ.* **610–611**, 1421–1427 (2018).
3. Shen, S. *et al. Lancet* **386**, S58 (2015).
4. Kinder, J. M. *et al. Cell* **162**, 505–515 (2015).

MEDICAL RESEARCH

Longevity data hint at no natural limit on lifespan

Death rates plateau in elderly people, reviving a debate about how long humans can live.

BY ELIE DOLGIN

There might be no natural limit to how long humans can live — at least not one yet in sight.

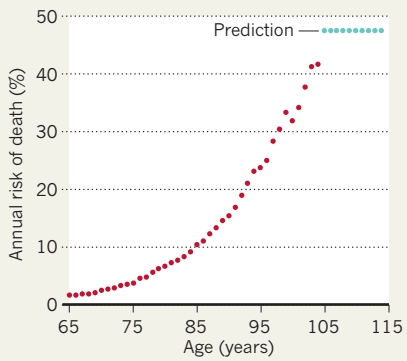
That proposal — which runs contrary to the claims of some demographers and biologists — comes from a statistical analysis published on 28 June in *Science*. It examined the probabilities of survival of nearly 4,000 'super-elderly' people in Italy, all aged 105 and older (E. Barbi *et al. Science* **360**, 1459–1461; 2018).

The study was led by Sapienza University demographer Elisabetta Barbi and University of Roma Tre statistician Francesco Lagona, both based in Rome. Their team found that the risk of death — which, throughout most of life, seems to increase as people age — levels off after age 105, creating a 'mortality plateau'. At that point, the researchers say, the odds of someone dying from one birthday to the next are roughly 50:50 (see 'Longevity unlimited').

"If there is a mortality plateau, then there is no limit to human longevity," says Jean-Marie Robine, a demographer at the French Institute of Health and Medical Research in Montpellier.

LONGEVITY UNLIMITED

A person's chances of dying tend to increase throughout adulthood, but a model based on data from 3,836 people aged 105 or older predicts that this trend flattens out in very elderly people.



SOURCE: BARBI *ET AL.* (2018).

That would mean that someone such as Chiyo Miyako, a Japanese great-great-great-grandmother who, at 117, is the world's oldest known person, could live for years to come — or even forever, at least hypothetically.

Researchers have long debated whether humans have an upper age limit. The consensus holds that the risk of death steadily increases in adulthood, up to about age 80 or so. But there's vehement disagreement about what happens as people enter their 90s and 100s.

Some scientists have examined demographic data and concluded that there is a fixed, natural 'shelf life' for our species, and that mortality rates keep increasing. Others have looked at the same data and concluded that the death risk flattens out in one's ultra-golden years, and therefore that human lifespan does not have an upper threshold.

In 2016, geneticist Jan Vijg and his colleagues at Albert Einstein College of Medicine in New York City rekindled the debate when they analysed the reported ages at death for the

© 2018 Springer Nature Limited. All rights reserved.

world's oldest individuals over half a century. They estimated that human longevity hit a ceiling at about 115 years — 125 tops.

Vijg and his team argued that given few, if any, gains in maximum lifespan since the mid-1990s, human ageing had reached its natural limit (X. Dong *et al. Nature* **538**, 257–259; 2016). The longest known lifespan belonged to Jeanne Calment, a French super-centenarian who died in 1997 at age 122.

Experts challenged the statistical methods in the 2016 study, setting off a firestorm into which Barbi and Lagona now step. Working with colleagues at the Italian National Institute of Statistics, the researchers collected records on every Italian aged 105 years and older between 2009 and 2015 — gathering certificates of death, birth and survival in an effort to minimize the chances of 'age exaggeration', a common problem among the oldest old.

They also tracked individual survival trajectories from one year to the next, rather than lumping people into age intervals as previous studies that combine data sets have done. And by focusing just on Italy, which has one of the highest rates of centenarians per capita in the world, they avoided the issue of variation in data collection between different jurisdictions.

As such, says Kenneth Howse, a health-policy researcher at the Oxford Institute of Population Ageing, UK, "these data provide the best evidence to date of extreme-age mortality plateaus in humans".

Ken Wachter, a mathematical demographer at the University of California, Berkeley, and an author of the latest study, suspects that previous disputes over the patterns of late-life mortality have largely stemmed from bad records and statistics. "If we can get data of this quality for other countries, I expect we're going to see much the same pattern."



ANTONIO CALANNI/AP/REX/SHUTTERSTOCK

Emma Morano, who died in 2017 at age 117, was the last surviving person born in the nineteenth century.

Robine is not so sure. He says that unpublished data from France, Japan and Canada suggest that evidence for a mortality plateau is "not as clear cut". A global analysis is still needed to determine whether the findings from Italy reflect a universal feature of human ageing, he says. Brandon Milholland, a co-author of the 2016 *Nature* paper, says that the evidence for a mortality plateau is "marginal", because the latest study included fewer than 100 people who lived to 110 or beyond. Leonid Gavrilov, a longevity researcher at the University of Chicago in Illinois, notes that even small inaccuracies in the Italian longevity records could lead to a spurious conclusion.

Others say the conclusions of the study are biologically implausible. "You run into basic limitations imposed by body design," says Jay Olshansky, a bio-demographer at the University of Illinois at Chicago, noting that cells that do not replicate, such as neurons, will continue to wither and die as a person ages, placing upper boundaries on humans' natural lifespan.

This study is thus unlikely to be the last word on the age-limit dispute, says Haim Cohen, a molecular biologist at Bar-Ilan University in Ramat-Gan, Israel. "I'm sure that the debate is going to continue." ■

CLIMATE CHANGE

Cyprus asserts itself as hub for climate research

Proposed science institute will focus on the Mediterranean and Middle East.

BY ANITA MAKRI

The tiny island of Cyprus is reshaping itself into a regional hub for climate-change research. The country lies at the meeting point of the Mediterranean, the Middle East and North Africa — areas where climate change is expected to take a heavy toll in the coming decades, but in which research capacity to address the issue is limited.

Cyprus's President Nicos Anastasiades announced plans on 5 June to create a government initiative that will coordinate action against global warming across the Mediterranean and support the creation of a €30-million (US$35-million) climate-change research centre at the Cyprus Institute in Nicosia, the nation's leading multidisciplinary research institution. "This is a priority issue for the government," says Theodoulos Mesimeris, head of the climate-change division of the Cypriot environment ministry. The initiative will also create a comprehensive plan for reducing Cyprus's greenhouse-gas emissions in line with goals set by the 2015 Paris climate accord.

Resources for climate research in the region are too small to scope out even the challenges, let alone the solutions, says Costas Papanicolas, president of the Cyprus Institute, who helped to plan the initiative with government ministers and Anastasiades.

Climate models suggest[1] that the Mediterranean and Middle East are getting warmer and drier at a rate faster than the global average; precipitation in the Mediterranean is expected to drop, especially in summer, by as much as 30–40% by the end of the century if no mitigation efforts are made, according to Filippo Giorgi, an Earth-systems physicist at the ▶

© 2018 Springer Nature Limited. All rights reserved.



PETROS KARADJIAS/AP/REX/SHUTTERSTOCK

Water scarcity is a growing problem in Cyprus and the surrounding region.

International Centre for Theoretical Physics in Trieste, Italy. Rains — when they come — will be more intense. Crop failures, forest fires and freshwater shortages[2] are just some of the issues that threaten economies, lifestyle and tourism. Parts of the region are set to become uninhabitable. In the Middle East, for instance, average maximum temperatures could increase from 43 °C to almost 50 °C by the end of this century, without mitigation[3].

"There is warming, and there is no mechanism to counteract the warming," says Jos Lelieveld, an atmospheric chemist at the Max Planck Institute for Chemistry in Mainz, Germany, who also works at the Cyprus Institute.

REGIONAL IMPACT

Few monitoring systems exist in the eastern Mediterranean and Middle East to systematically measure variables such as temperature, humidity and desertification. The monitoring that does exist is inconsistent, and the data are too poor to feed into climate-change models, which would help researchers to understand local impacts and refine policy options.

At the core of the proposed hub — the Eastern Mediterranean Middle East Climate and Atmosphere Research Centre — will be a high-quality observatory for monitoring concentrations of greenhouse-gas emissions and atmospheric contaminants, which will take advantage of Cyprus's geographical location to establish the region's contributions.

The centre will absorb the existing climate-research activities of the Cyprus Institute. The institute, launched in 2007, has already helped to raise awareness of the issue in the region, says Khaled Toukan, chairman of the Jordan Atomic Energy Commission and the country's former energy minister. Jordan and other countries in the Middle East are moving towards clean energy, he says, but purely from an economic perspective.

Papanicolas says that the institute is capitalizing on Cyprus's position as the only European Union country in the Middle East. It has already won €400,000 in EU research money to develop a plan for the facility, and it is now preparing a bid for €15 million in EU funding, which would be matched by the Cypriot government and would bankroll the centre for the next decade. ■

1. IPCC. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects* (Cambridge Univ. Press, 2014).
2. Lelieveld, J. *et al. Clim. Change* **114**, 667–687 (2012).
3. Lelieveld, J. *et al. Clim. Change* **137**, 245–260 (2016).

ASTRONOMY

NASA telescope's woes grow

The James Webb Space Telescope's cost and schedule problems threaten other big missions.

BY ALEXANDRA WITZE

NASA's beleaguered James Webb Space Telescope (JWST) is facing yet another delay, and will not launch until March 2021. That's ten months later than the tentative schedule that the agency announced just three months ago. To meet the new target, NASA must persuade lawmakers in Congress to approve a higher price for the mission.

The space agency estimates that the latest delay will add US$800 million to the telescope's cost, on top of the $8 billion Congress has already approved for its development. NASA plans to make up that shortfall in part by using money that had been intended to support the telescope's science operations in space. Still, the delays will loom over the agency's astrophysics budget, with unknown effects on the next big space telescope in NASA's queue: the Wide-Field Infrared Survey Telescope (WFIRST).

"I'm not happy sitting here," said Thomas Zurbuchen, NASA's associate administrator for science, at a 27 June news briefing on the delay. But he said that ensuring a successful mission was worth the extra time and money. Among other things, the telescope will peer back in time to explore some of the earliest galaxies to form in the Universe, and will probe the atmospheres of planets around other stars.

"JWST should continue because of the compelling science."

"JWST should continue because of the compelling science and because of its national importance," said Thomas Young, a retired executive with Lockheed Martin in Bethesda, Maryland. He oversaw an independent review of the telescope project that led to the revised schedule and budget estimates.

Members of Congress have sharply criticized NASA for previous JWST delays, and the latest announcement has continued the pattern. "Programme delays and cost overruns don't just delay the JWST's critical work, but they also harm other valuable NASA missions, which may be delayed, defunded, or discarded entirely," said Representative Lamar Smith (Republican, Texas), the chairman of the House science committee, in a statement.

JWST is the most complex astronomical telescope ever built, and problems have piled up towards the end of its development. The observatory is undergoing extensive testing at Northrop Grumman Aerospace Systems in Redondo Beach, California. The independent review

© 2018 Springer Nature Limited. All rights reserved.

found that engineers made several errors at Northrop, including using the wrong solvent to clean valves, which later leaked, and not tightening the sunshield fasteners properly.

The JWST has a 6.5-metre-wide segmented mirror that will launch in a folded configuration and then unfurl once it is in space. The telescope's sunshield must also deploy without a hitch. JWST is technologically more complex than the Hubble Space Telescope, whose primary mirror was ground incorrectly, a problem discovered after launch. Astronauts fixed Hubble's vision in low Earth orbit, but repairs in space won't be possible for JWST, which will orbit 1.5 million kilometres from Earth.

Until last September, JWST was on track for an October 2018 launch. Then NASA pushed the date to June 2019, then May 2020 — and now March 2021.

DECADAL DETOUR

The previous delay, which NASA announced in March, had prompted Zurbuchen to propose that the US astronomy community postpone its next 'decadal survey'. This influential process, which takes place every ten years, asks astronomers to decide which scientific questions their field should tackle and what facilities they need to answer those questions. Preparations for the next survey, which is due in 2020, were well under way when Zurbuchen suggested delaying it.

But the two other agencies involved in the survey — the US National Science Foundation and the Department of Energy — did not want to put it off. And neither did most of the astronomers polled by the US National Academies of Sciences, Engineering and Medicine, which oversees the survey. In late May, Zurbuchen reversed his stance, and the survey is now on track to meet its original deadline.

On 27 June, Paul Hertz, NASA's head of astrophysics, told an advisory committee that the rising costs for JWST were "likely to impact other science programmes". At particular risk is WFIRST, slated for launch in the mid-2020s. Just as JWST was the highest priority recommendation from the 2000 decadal survey, WFIRST was the highest priority in the 2010 decadal survey. Both are being developed by NASA's astrophysics division. The Trump administration has proposed cancelling WFIRST, although Congress has so far come to its rescue with continued funding.

In the long run, JWST promises revolutionary research that cannot be achieved any other way, says Jason Kalirai, an astronomer at the Space Telescope Science Institute in Baltimore, Maryland, and the project scientist for the telescope. "We need JWST to make the next big breakthroughs in astrophysics — and are willing to wait for it," he says. ■

FUNDING

EU crackdown on 'ethics dumping'

Fund aims to stop scientists exporting dubious research.

BY LINDA NORDLING

Ethics dumping — doing research deemed unethical in a scientist's home country in a foreign setting with laxer ethical rules — will be rooted out in research funded by the European Union, officials announced last week.

Applications to the EU's €80-billion (US$93-billion) Horizon 2020 research fund will face fresh levels of scrutiny to make sure that research practices regarded as unethical in Europe are not exported to other parts of the world. Wolfgang Burtscher, the European Commission's deputy director-general for research, made the announcement at the European Parliament in Brussels on 29 June.

Burtscher said that a new code of conduct developed to curb ethics dumping will soon be applied to all EU-funded research projects. That means applicants will be referred to the code when they submit their proposals, and ethics committees will use the document when considering grant applications.

The rules will apply to all research funded under Horizon 2020, and to all future EU funding programmes. The EU had banned ethics dumping in Horizon 2020 grants since 2013. But no clear guidelines existed to help ethics reviewers and researchers identify potential digressions in grant applications. The code, drafted as part of a Horizon 2020-funded project called TRUST, was published in May; the latest announcement gives it teeth.

The code provides clear guidance for doing research in resource-poor settings. Animal studies, for example, must not be conducted outside the EU if they would not be allowed in the scientists' home country. Another provision states that "lower educational standards, illiteracy or language barriers" among research participants can never be an excuse to hide information from them or provide it incompletely. The code also addresses situations that might not arise in Europe-based studies. For instance, sex work is legal in many countries in Europe but not in Kenya. And homosexuality is illegal in many countries worldwide. So studies involving sex workers or gay people, for example, must take measures to ensure the safety of participants.

The ethics-dumping guidelines were produced with representatives from such vulnerable populations. Joyce Adhiambo, a Kenyan former sex worker who promotes sex workers' rights in research and in HIV-prevention services, sees the code as a matter of mutual respect. "When [researchers] want something from sex workers, we deal with it respectfully. We ask the same in return," she said at the Brussels event.

Adhiambo told *Nature* that researchers must use their privileged position to encourage communities to become actively involved in studies. Members could be hired as research assistants, for example, or to help translate and explain consent forms to participants. "We come from a poor setting but we have a voice. We have a culture and a way of living. We have our traditional knowledge, and when we walk in the path together, we are going to make a brighter future for all these research projects."

Ethics dumping — coined by the European Commission in 2013 — is a contentious term, and few researchers admit to the practice. In a recent book, researchers with the TRUST project cited research carried out on wild-caught monkeys in Africa, and clinical trials in India in which people living in poverty were denied life-saving screening in the control arm, as examples of ethics dumping (see go.nature.com/2mkhcx4). ▶

MORE ONLINE

STAY CURRENT

- Pollution responsible for one-fifth of infant deaths in sub-Saharan Africa go.nature.com/2il7ckr
- Mysterious interstellar visitor is a comet go.nature.com/2nnqnej
- Australia moves a step closer to 'three-person IVF' go.nature.com/2kligdy

NATURE PODCAST

A DNA computer; the koala genome; and the invisibility of LGBTQ+ researchers nature.com/nature/podcast

© 2018 Springer Nature Limited. All rights reserved.

SIEGFRIED MODOLA/REUTERS



A research code of ethics contains input from vulnerable groups, including sex workers in Kenya.

None of those projects was funded by the EU, says Doris Schroeder, a lead investigator on the TRUST project. But in her 15 years chairing ethics review panels for EU funding programmes, Schroeder has seen many applications that would violate the new code. These ranged from researchers wanting to interview workers about their rights in dictatorial states (potentially placing these people at risk) to art installations portraying vulnerable populations without their involvement. Those projects were changed before getting funding approval, Schroeder says. But without a clear code of conduct, it's possible that other ethics committees might have let them through.

Ron Iphofen, an adviser on research ethics to the European Commission, thinks that the code will have a profound impact on how funding proposals to the EU are designed and reviewed. "I could envisage reviewers now looking suspiciously at any application for funds that entailed research by wealthy nations on the less wealthy that did not mention the code," he says.

Opportunities for ethics dumping have grown with the globalization of research, says Philip Brey, a research-ethics specialist at the University of Twente in the Netherlands. Increasingly, researchers from high-income countries carry out projects in low- and middle-income ones. But Brey says that the decision to export research is often driven by scientific opportunities or economic realities, rather than by a desire to skirt ethics. Moreover, some scientists in poorer countries find the term 'ethics dumping' offensive. "They tend to see themselves not as having lower ethical standards, but different ethical standards," says Brey.

Reinhard Hiller, managing director of the Centre for Proteomic and Genomic Research in Cape Town, South Africa, worries that, in some cases, developed nations' ethical standards could stifle research in developing nations. For example, to speed up and improve the quality of their diagnoses, doctors in Africa might want to use WhatsApp to share patient information such as X-rays, says Hiller. Yet this could fall foul of Europe's strict data-privacy rules, for example. "It's not black or white, but needs to be assessed on a case-by-case basis," he says. ■

© 2018 Springer Nature Limited. All rights reserved.



Embryo assembly 101

Researchers are starting to demystify the earliest stages of human development — edging right up to an ethical red line.

BY HELEN SHEN

For two tense weeks in mid-2013, developmental biologist Magdalena Zernicka-Goetz was chasing a world record. She and her colleagues at the University of Cambridge, UK, were attempting to grow human embryos in the lab for longer than had ever been done before. They wanted to glean insights into how a tiny blob of cells transforms itself into a complex, multipart structure. Previous efforts had stalled after about a week, but Zernicka-Goetz knew there was much more to learn about human development beyond that point.

The researchers started with embryos that had been donated by women who no longer needed them for *in vitro* fertilization (IVF) procedures. The team bathed the cells in a special medium and housed them in an incubator, using methods adapted from their previous work on mouse embryos. Because the samples had to stay in a strictly controlled environment, the scientists could remove them only once or twice a day to track their progress under a microscope.

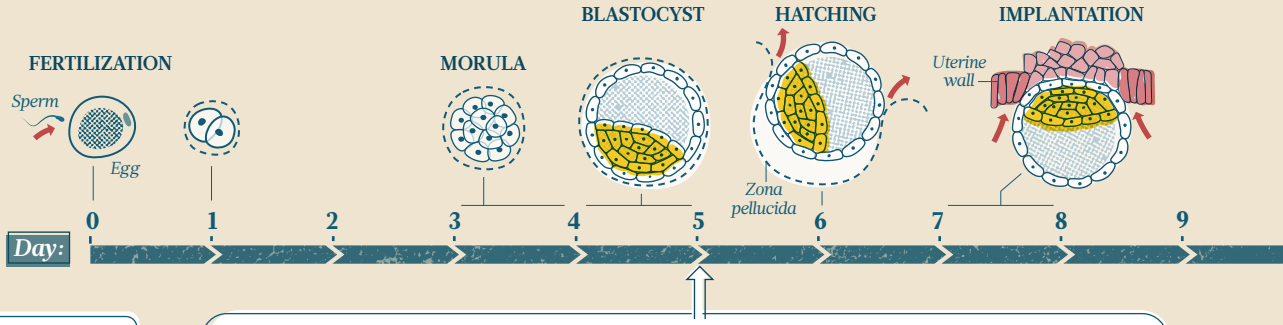
The days ticked by — six, seven, eight. And still, the embryos continued to thrive and develop[1]. "We would hold our breath," recalls Zernicka-Goetz. "Each day was more and more exciting." The team reached 12 days on its first attempt, eventually stretching to 13. "This was just unbelievable. I was so delighted," she says.

Their advance, and a similar feat by a group based in New York City[2], is one of a few achievements in the past five years to heat up the study of early human development. Researchers'

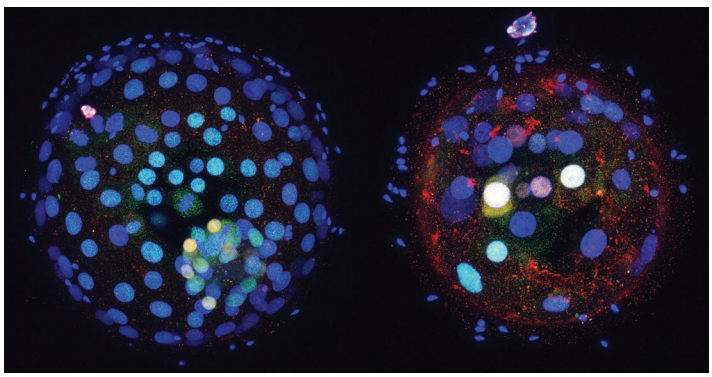
ILLUSTRATION: TOBY LEIGH

© 2018 Springer Nature Limited. All rights reserved.

Baby steps



For decades, researchers could only guess at the early stages of human development, using animal studies and rare tissue samples as a guide. New methods for nurturing whole human embryos in the lab — and building embryo-like constructs from human stem cells — are starting to crack open the black box.



Using the gene-editing technique CRISPR, researchers blocked a key protein in early development called OCT4 and watched human embryos fail to grow into 200-cell blastocysts. Mouse embryos lacking OCT4 faltered at a later stage, hinting at differences between the two species even very early in development.

NIK SPENCER/*NATURE*; IMAGES: LEFT TO RIGHT: NORAH FOGARTY AND KATHY NIAKAN; MAGDALENA ZERNICKA-GOETZ GROUP; I. MARTYN *ET AL./NATURE*

access to the human embryo has always been limited, and they knew relatively little about its early transformations. But now, refinements in cell-culturing methods are enabling them to grow human embryos outside of the body for up to two weeks. Scientists are using gene-editing techniques, such as CRISPR, and building artificial embryo-like structures to explore the cellular signals and physical forces that shape the embryo and its supporting cast of tissues.

These techniques are illuminating key early processes, such as implantation — when the minuscule embryo embeds itself into the uterine wall and can't be studied directly. And new high-resolution, digital images are revealing in fine detail how muscles and nerves grow a few weeks later in development. Such discoveries could lead to a better understanding of how birth defects and developmental disorders arise, as well as why some pregnancies fail.

But alongside their promise, these new techniques are pushing researchers into uncharted ethical territory. Beginning in the late 1970s, ethicists and scientists converged on the '14-day rule', which limits work on human embryos to a fortnight after fertilization — a time when the first hints of the nervous system appear, and the last point at which an embryo can divide. Until now, the internationally recognized 14-day rule has been a purely hypothetical limit. "It wasn't a rule that anybody was butting up against," says bioethicist Josephine Johnston at the Hastings Center in Garrison, New York. "It's now technically possible."

INSTRUCTION MANUAL

Many early developmental processes are surprisingly similar throughout the animal kingdom, with each species tweaking a few genes here or signals there. Among mammals, scientists have studied the mouse molecular instruction manual the most, disabling genes one by one to test what they do. Mice are easy to obtain in the numbers often needed for experiments, and are considered a decent proxy for studying human embryonic development — many of the earliest cell types and components seem similar in both species. But researchers are starting to question how far these similarities really go. "As we've been able to explore early human development a little bit, it's become apparent the mouse and human embryo are similar, but not the same," says developmental biologist Janet Rossant at the Hospital for Sick Children in Toronto, Canada.

With a limited supply of human tissue available, scientists have turned to highly efficient gene-editing technologies such as CRISPR–Cas9 to explore the early stages of embryo development. In part owing to ethical sensitivities surrounding genetic modification of embryos, only a few groups have received authorization to perform such studies so far.

At the Francis Crick Institute in London, developmental biologist Kathy Niakan led the first project of its kind to receive approval from national regulators. In 2017, her team reported using CRISPR–Cas9 to edit a gene expressed in both human and mouse embryonic stem cells[3]. Human embryos with disruptions to this gene lacked a protein called OCT4 and failed to develop into blastocysts — balls of roughly 200 cells. By contrast, mouse embryos lacking the same gene formed blastocysts and faltered only later.

The difference supports the growing idea that, even in very early development, some genetic details — such as when certain genes are active — might be specific to humans. "We know that most IVF embryos will fail to develop to the blastocyst stage," says Niakan. The reasons for that are unclear. "Understanding which pathways might be responsible for specifying the first cell types in the embryo could potentially lead to improvements in IVF." In future work, Niakan hopes to examine the human genes that commit a tiny fraction of blastocyst cells to forming the embryo proper, as opposed to supporting tissues such as the placenta.

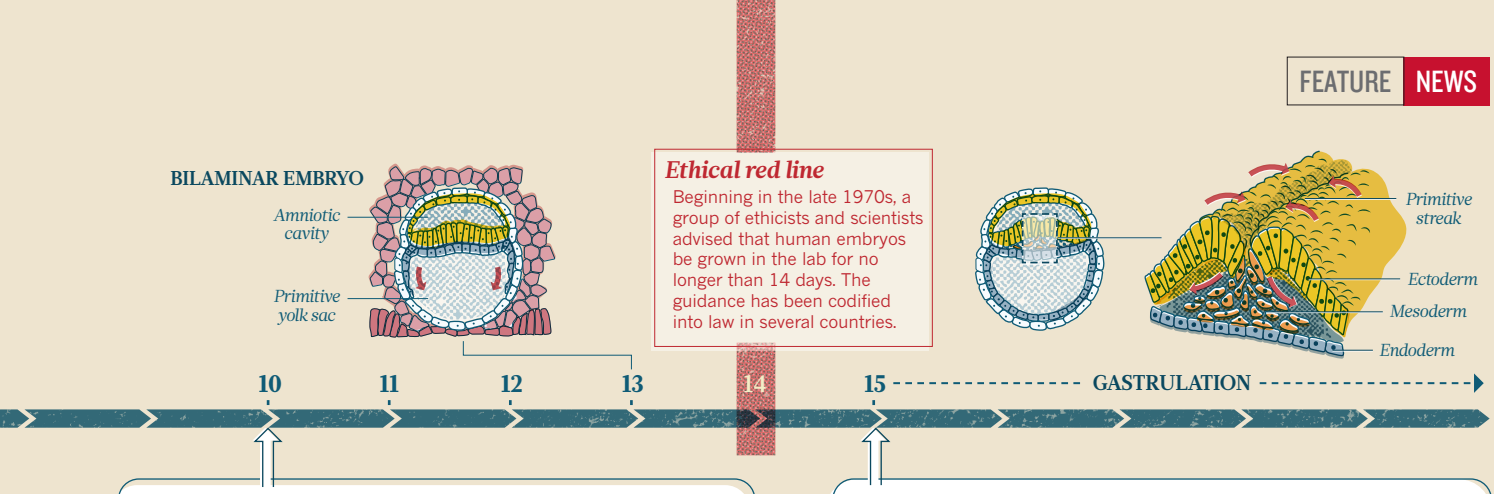
FINDING A HOME

After blossoming into a 200-cell ball, the little blastocyst must embed into the uterine wall to survive. But once this happens (around day seven), scientists are largely unable to study its development. Observing the implantation process itself is the first challenge: until recently, researchers lacked reliable methods for sustaining embryo development beyond the first week.

Now, scientists have opened that black box. In two papers published in May 2016, Zernicka-Goetz's team[1] and Ali Brivanlou's group[2] at the Rockefeller University in New York City reported the first culture systems that could grow human embryos for 12–13 days. The researchers showed that with the right cocktail of growth factors and nourishment, human embryos in culture can 'implant' onto the bottom of the dish. Remarkably, the embryos didn't require any maternal tissue to trigger the early remodelling steps that occur after implantation. "That was shocking to me," recalls Brivanlou. "I would have thought it impossible for the human embryo to even go past one or two days after attachment."

Embryos that latch onto the dish are flatter than the real thing. (Brivanlou likens the

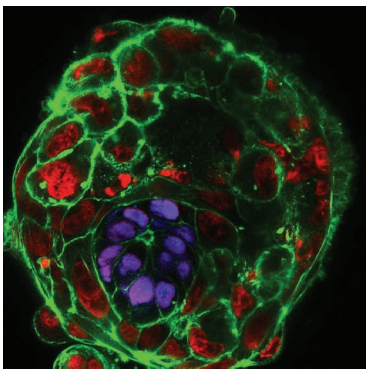
© 2018 Springer Nature Limited. All rights reserved.



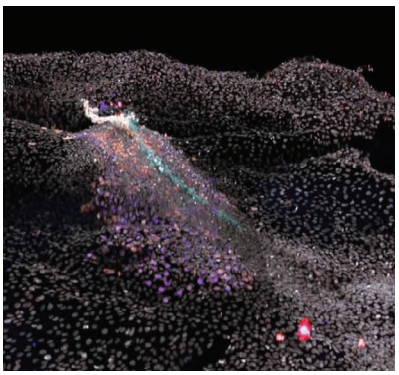
Ethical red line

Beginning in the late 1970s, a group of ethicists and scientists advised that human embryos be grown in the lab for no longer than 14 days. The guidance has been codified into law in several countries.

Remarkably, embryos can begin to self-organize without implanting in a womb. Cells in this lab-grown embryo have begun to differentiate into types — purple cells here will become the embryo proper — and have started to form the amniotic cavity, which will enclose the embryo as it grows.



In animals, 'organizer cells' choreograph the formation of the embryo's head-to-tail axis and the beginnings of its nervous system. Working with synthetic embryos made from human stem cells, researchers recently demonstrated for the first time the existence of these organizers in humans.



attachment process to a parachute landing.) But the cultured embryos nevertheless went on to hit several milestones expected from animal experiments and from limited studies of human tissue samples collected after miscarriages and from other sources. In the latest experiments, after the embryo attached to the dish, an outer cell layer began differentiating into early placental and other cell types that support embryonic growth. Internally, cells seemed to develop into precursors of the embryo proper and yolk sac — an early structure supplying blood to the embryo. After almost a fortnight was up, both teams ended the experiments, in accordance with the 14-day rule.

By then, some of the embryos had stalled in their development, but with further improvements, the researchers say, these culture systems could help to uncover much more detail about the embryo's first two weeks.

Several key events happen in the third week, during a process called gastrulation. The embryo starts to develop a body axis with one end destined to become the head. And cells start to migrate and differentiate into the three layers that will eventually produce all of the body's organs and tissues (see 'Baby steps'). The importance of this process was a major reason for imposing an ethical limit of 14 days.

But some researchers are finding alternative approaches, using human stem-cell technology to construct synthetic embryo-like structures, which are not covered by the 14-day rule. These constructs lack certain components essential for full development, and couldn't give rise to a human if implanted. In 2014, Brivanlou, Eric Siggia and their colleagues at the Rockefeller University reported mimicking gastrulation *in vitro* using specially grown human embryonic stem cells[4]. They found that when stem cells are confined to grow in circles measuring a few hundred micrometres across, they differentiate into a bulls-eye pattern containing the three main cell types that give rise to all parts of the future body. In nearly all animals, from flatworms to primates, these cell types have been found to have a similar role: cells in the centre become skin, brain and nervous system; in the next ring, muscle, blood, bones and various organs; and in the outer ring, the digestive tract and respiratory system.

The flattened rings don't look like the 3D

> ***"Here's a system where we can dissect relationships between signalling pathways and cell fates."***

triple-decker sandwich that forms in real human embryos, but at a cellular and molecular level, the construct did just what researchers expected. "Here's a system where we can really dissect relationships between signalling pathways and cell fates," says Aryeh Warmflash, a former postdoctoral fellow in Siggia's lab.

Follow-up studies of the system have already revealed new details about how embryonic cells use geometry and chemistry to self-organize into distinct tissue types. In 2016, Brivanlou, Siggia and their teams demonstrated[5] that stem cells can sense their position in the circular colonies and, accordingly, adjust how they respond to molecules called growth factors, helping to form discrete zones of cells. And in a preprint posted to the bioRxiv server in 2017, Warmflash's group at Rice University in Houston, Texas, demonstrated that the dynamics of growth factors have a role, too; the researchers saw that a surge in signalling by a particular set of proteins — collectively called the Nodal pathway — spreads from the colony perimeter inwards like a wave, leaving different cell types in its wake[6].

Brivanlou's group has taken the system even further, showing in May[7] that treating these types of colonies with a combination of growth factors induces the formation of 'organizer' cells. In animals, these special cells direct their neighbours to form a head-to-tail axis. But, in part because of the 14-day rule, scientists had never seen human organizer cells in action. Given the ethical and technical limitations of working with human embryos, Brivanlou instead grafted clusters of putative human organizer cells onto developing chick embryos, and watched as the grafted cells directed chick cells to develop into a second chicken nervous system.

LIFE SUPPORT

Much research on early human development has focused on the embryo itself, but many other tissues are crucial to its survival. These include the amniotic sac, which houses the embryo, and the placenta, which provides oxygen and nutrients.

To better study how the amniotic sac develops, researchers created a model using human stem cells. Last year[8], developmental biologist Deborah Gumucio and bioengineer Jianping Fu at the University of Michigan in

© 2018 Springer Nature Limited. All rights reserved.

Ann Arbor and their colleagues showed that when they grew human stem cells on a gel bed and surrounded them with natural scaffold molecules, the cells self-organized into a clump resembling the amniotic sac. After about 24 hours, a hole opened up and, subsequently, cells began to flatten on one side and elongate on the other, characteristic of the process leading up to gastrulation.

The model replicates only the core of the embryo and its sac, with none of the other supporting tissues it would need to be viable; however, the researchers were still able to identify some of the molecular signals that could help to give rise to this asymmetrical structure. "There's so much good information to be found, there's no reason to push this in any way to get close to making embryos," says Gumucio. The scientists maintained their embryo-like cultures for up to 5 days, corresponding roughly to developmental days 9–14, at which point they ended the experiments.

Zernicka-Goetz's team is trying to develop more-complete structures. In 2017, the team cultured a combination of two types of mouse stem cell: those that form the embryo itself, and some that help to form the placenta, called trophoblast stem cells. Embedded in a 3D scaffold, this synthetic structure grew to resemble the embryo after implantation[9]. Now, the researchers are working on creating a similar embryo-like construct using human stem cells. Such a development could help scientists to learn more about the cross-talk between embryonic tissues and extraembryonic ones, such as the placenta.

As some labs develop these increasingly sophisticated synthetic models, called embryoids, ethical questions are beginning to bubble up. "I think it really is a grey area," says Martin Pera, a stem-cell biologist at The Jackson Laboratory in Bar Harbor, Maine. "How do we regard these structures that are developing?" Many ethicists and scientists agree that current versions of synthetic embryoids are too simplistic to fall under the 14-day rule. But bioethicist Insoo Hyun at Case Western Reserve University in Cleveland, Ohio, says that it's a challenge to define which features would make embryoids just realistic enough. "The potential is there for something to be constructed that's much further along than 14 days, and that could develop if you were to implant it into the uterus."

> *"In terms of detail, there was nothing like that before."*

BODY BUILDING

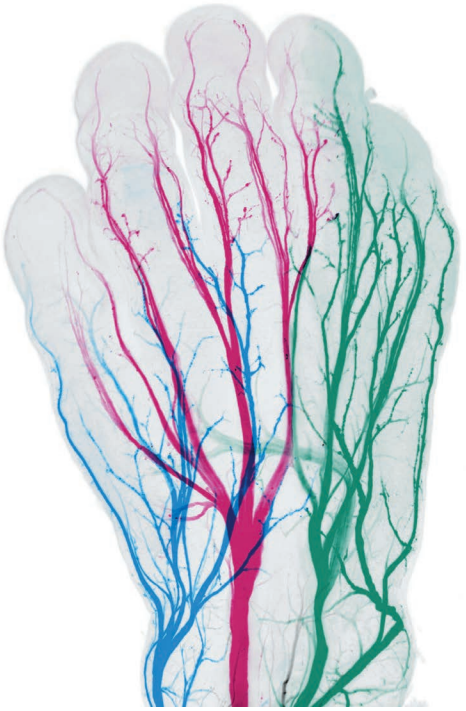
Researchers are also making strides in unlocking details from later phases. Many of these advances come from better tissue staining and imaging. In a 2016 report in *Science*, researchers at the University of Amsterdam's Academic Medical Center digitized 15,000 slices from a collection of tissues maintained by the Carnegie Institution for Science in Washington DC. The collection houses samples obtained from miscarriages, surgeries and autopsies, beginning in the 1880s and continuing over the first half of the twentieth century. In the Dutch study[10], which spans the first 2 months of development, the researchers digitally traced the outlines of up to 150 organs on individual slices, and then aligned the sections to reconstruct 3D models of the original embryo, available as an interactive atlas.

Studying the embryos in high resolution has already led to a few discoveries. For instance, the team found[10] that the kidneys, which are thought to ascend, and the gonads, which are thought to descend during development, only seem to rise and fall within the body because they grow at different rates compared to the vertebrae.

But one major limitation of the Carnegie collection is its lack of molecular markers, which makes it difficult to distinguish different cell types. To address these issues, Alain Chédotal at the French National Institute of Health and Medical Research (INSERM) in Paris published in 2017 a new 3D atlas of 36 human embryos and fetuses spanning weeks 6–14 of development[11]. Chédotal and his team applied a tissue-clearing treatment to the donated specimens to make the samples easier to image under a microscope, and they stained them to highlight various cell types.

The resulting 3D images show, in high-resolution detail, developing nerves, muscles, lungs, and other organs. "In terms of detail, there was nothing like that before," says Rui Diogo at Howard University College of Medicine in Washington DC, who has mined the data set for information on how limb muscles develop.

In still unpublished results, Diogo's team has observed a number of hand and foot muscles that disappear or fuse during development. "When we are embryos, we have muscles we don't have as adults," he says. Among other unexpected findings, Chédotal reported that, although the gross organization of nerves is similar in the left and right hands, the fine branching patterns in each hand diverge and strike out along different paths in the first 7–11 weeks.



ALAIN CHÉDOTAL/MORGANE BELLE/FONDATION VOIR ET ENTENDRE

This image shows the nerves branching in a hand at nine weeks of gestation.

DEVELOPING FIELD

As scientists refine and improve their techniques, they hope to learn much more about human development, and to shed light on the causes of pregnancy loss and birth defects.

As an embryo's earliest days come into sharper focus, so too do discussions of the ethical limits. Zernicka-Goetz's and Brivanlou's extended experiments have led some to suggest that the time is ripe to re-evaluate the pros and cons of the 14-day rule. In May, the Baker Institute Center for Health and Biosciences at Rice University hosted a meeting of 30 US scientists, ethicists and other experts, including Brivanlou and Johnston, to discuss whether and how to move the boundary. "I think it's better to keep the 14-day rule in place and have a special petition to make an exception," says Hyun, who also attended the event.

As the results of this research accumulate, the technical advances are inspiring a mixture of fascination and unease among scientists. Both are valuable reactions, says Johnston. "That feeling of wonder and awe reminds us that this is the earliest version of human beings and that's why so many people have moral misgivings," she says. "It reminds us that this is not just a couple of cells in a dish." ■

Helen Shen *is a science journalist based in Sunnyvale, California.*

1. Shahbazi, M. N. *et al. Nature Cell Biol.* **18**, 700–708 (2016).
2. Deglincerti, A. *et al. Nature* **533**, 251–254 (2016).
3. Fogarty, N. M. E. *et al. Nature* **550**, 67–73 (2017).
4. Warmflash, A., Sorre, B., Etoc, F., Siggia, E. D. & Brivanlou, A. H. *Nature Meth.* **11**, 847–854 (2014).
5. Etoc, F. *et al. Dev. Cell* **39**, 302–315 (2016).
6. Heemskerk, I. *et al.* preprint at bioRxiv https://doi.org/10.1101/202366 (2017).
7. Martyn, I., Kanno, T. Y., Ruzo, A., Siggia, E. D. & Brivanlou, A. H. *Nature* **558**, 132–135 (2018).
8. Shao, Y. *et al. Nature Commun.* **8**, 208 (2017).
9. Harrison, S. E., Sozen, B., Christodoulou, N., Kyprianou, C. & Zernicka-Goetz, M. *Science* **356**, eaal1810 (2017).
10. de Bakker, B. S. *et al. Science* **354**, aag0053 (2016).
11. Belle, M. *et al. Cell* **169**, 161–173.e12 (2017).

© 2018 Springer Nature Limited. All rights reserved.

COMMENT

EQUALITY Too few people who are LGBTQ go into science, and too many leave **p.27**

HERITAGE How UNESCO has tried to broker peace through science and culture **p.29**

EVOLUTION Synthesizing many lines of evidence to trace the spine's start **p.31**



BIOLOGY Wet labs squeezed by scarce funds and bureaucracy in Italy **p.32**

ILLUSTRATION BY DAVID PARKINS



A better measure of research from the global south

Funders **Jean Lebel** and **Robert McLean** describe a new tool for judging the value and validity of science that attempts to improve lives.

In India, the world's leading producer of mangoes, up to 40% of the harvested fruit is destroyed in transit before delivery. This costs up to US$1 billion in lost income each year, affecting the lives and livelihoods of millions of farmers, traders and consumers. So researchers from India, Sri Lanka and Canada developed a suite of nanomaterials that can be sprayed onto fruit on the tree, in packaging or in transit, to extend its life. They trapped hydrophobic hexanal molecules (derived from plant waste) in a hydrophilic membrane so that they could be suspended in liquid for application to the fragile fruit.

In Egypt, more than 95% of women have experienced sexual harassment at least once, and most cases go unreported. So, in 2010, researchers at the Youth and Development Consultancy Institute in Cairo developed Harrassmap (https://harassmap.org/en). This online interactive resource enables people to report and map cases of sexual harassment. When it emerged that university campuses were hotspots, Cairo University implemented a policy to combat sexual harassment, the first of its kind in the Middle East. Other universities in Egypt are following suit.

Both projects help to solve pressing societal challenges. The researchers involved ▶

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

appreciate that the people who benefit from the projects are the ones who are best placed to judge the value and validity of the work. The research teams spent time developing their hypotheses and results with those who feel the effects. In each case, the research is robust and life-changing — exactly the combination that most people would say is the very purpose of science.

But both projects would score poorly if judged using only conventional approaches to evaluating research quality that prioritize the opinion of peers, the volume of papers published, and citations. That's a problem because it is endorsement from other scientists, not stakeholders, that drives career advancement for researchers in Egypt, Sri Lanka and India, as everywhere else.

Is the weakness in the science or in the way it is measured? Too often it is the latter, in our view. Dominant techniques of research evaluation take a narrow view of what constitutes quality, thus undervaluing unique solutions to unique problems. At Canada's International Development Research Centre (IDRC) in Ottawa, we fund just this sort of research: natural and social science that unearths fixes for the development challenges facing countries in the global south. The majority of the work we support is led by researchers from these countries.

So we at the IDRC developed a tool to evaluate the quality of research that is grounded in, and applicable to, the local experience. We used it to assess 170 studies and then did a meta-analysis of our evaluations. The results suggest that it is possible — and essential — to change how we assess applied and translational research.

TUNNEL VISION

The limitations of dominant research-evaluation approaches are well known[1–5]. Peer review is by definition an opinion. Ways of measuring citations — both scholarly and social — tell us about the popularity of published research. They don't speak directly to its rigour, originality or usefulness. Such metrics tell us little or nothing about how to improve science and its stewardship. This is a challenge for researchers the world over.

The challenge is compounded for researchers in countries in the global south. For instance, the pressure to publish in high-impact journals is a steeper barrier because those journals are predominantly in English and biased towards publishing data from the United States and Western Europe[6]. With the exception of an emerging body of Chinese journals, local-language publications are broadly deemed lower tier — even those published in European-origin languages such as Spanish, Portuguese or French.

The metrics problem is further amplified for researchers who work on local challenges. Climate adaptation research is a case in point. Countries in the global south are on the front lines of global warming, where context-appropriate adaptation strategies are crucial. These depend on highly localized data on complex factors such as weather patterns, biodiversity, community perspectives and political appetite. These data can be collected, curated, analysed and published by local researchers. In some cases, it is crucial that the work is done by them. They speak the necessary languages, understand customs and culture, are respected and trusted in communities and can thus access the traditional knowledge required to interpret historical change. This work helps to craft adaptations that make a real difference to people's lives. But it is also fundamental to high-level meta-research and analysis that is conducted later, far from the affected areas[7].

Does the current evaluation approach scrutinize and give equal recognition to the local researcher who focuses on specifics and the researcher who generalizes from afar? Does the current approach acknowledge that incentives are different for local and foreign researchers, and that those incentives affect research decisions? Are we adequately measuring and rewarding research that is locally grounded and globally relevant? In our view, the answer to all of these questions is no.

FROM NO TO YES

With the support and leadership of partners across the global south, the IDRC decided to try something different. The result is a practical tool that we call Research Quality Plus (RQ+)[8].

The tool recognizes that scientific merit is necessary, but not sufficient. It acknowledges the crucial role of stakeholders and users in determining whether research is salient and legitimate. It focuses attention on how well scientists position their research for use, given the mounting understanding that uptake and influence begins during the research process, not only afterwards.

"Conventional evaluations were never this challenging, but neither were they so motivating."

We think that the approach has merit beyond the development context. We hope that it can be tailored, tested and improved in a variety of disciplines and contexts, to suit the needs of other evaluators — funders such as ourselves, but also governments, think tanks, journals and universities, among others.

RQ+ has three tenets:

Identify contextual factors. There is much to learn from the environment in which research occurs. Instead of aiming to isolate research from how, where and why it was done, and by whom, evaluators should examine these contexts to reach a claim about quality. For the IDRC, this included five issues: political, data, research environments, the maturity of the scientific field and the degree to which a project includes a focus on capacity strengthening. For another funder, journal or think tank, these might — or should — be different.

Articulate dimensions of quality. The underlying values and objectives of the research effort need to be made explicit. Evaluators weigh these dimensions of quality using a formula that fits the context and goals of the research. The dimensions that matter to the IDRC are: scientific integrity (a measure of methodological rigour), legitimacy (a measure of the fidelity of the research to context and objectives), importance (a measure of relevance and originality) and positioning for use (the extent to which research is timely, actionable and well communicated). (See Figure S1 in Supplementary Information.)

Use rubrics and evidence. Assessments must be systematic, comparable and based on qualitative and quantitative empirical evidence, not just on the opinion of the evaluator — no matter how expert they are. For the IDRC, this meant evaluators speaking to intended users, to others working in similar areas and to non-scientific beneficiary communities, as well as assessing research outputs and associated metrics.

ROAD TEST

The IDRC first used RQ+ in 2015. Independent specialists assessed 170 studies from 7 areas of research the centre had funded in the previous 5 years. For each area, three specialists rated projects using the three tenets described, looking at empirical data for each study: bibliometrics, interviews with stakeholders and IDRC reports on the work. The reviewers decided independently what data to collect and compare for each project, and held panel discussions to reach a consensus on the final ratings for each project.

This framework (see Figure S2 in Supplementary Information) encouraged a grounded, critical reflection on each project. And it helped systematic judgement to be applied across diverse contexts, disciplines and approaches to research. In exit interviews and follow-up discussions, the independent reviewers described the assessments as unlike any others they had done. They felt confident that the evaluation had been systematic, comprehensive and fair.

We learnt a lot from this process about the projects that the IDRC supports and how we could do better. For instance, we found that we need to prioritize gender across everything we fund, from climate modelling to the accessibility of justice, and not just in research projects that are

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

CLIFF CHENEY/ZUMA WIRE



Women protest against sexual harassment in Cairo in 2013.

aimed specifically at women and girls. As enshrined in one of the United Nations Sustainable Development Goals (SDG5), gender equality is key for unlocking development potential, so it was a dimension examined by the reviewers.

They found, for example, that a programme using national data sets to examine the implications of taxation and food labelling should have disaggregated the data by gender to achieve more with the same investment. Reviewers also highlighted exemplars, such as the African Doctoral Dissertation Research Fellowship programme, which helps PhD students to complete theses at their home institutions, enabling greater uptake by female applicants who shoulder more family duties. The programme considers gender balance when selecting applicants, and in reviewing proposed research.

As a result, the IDRC has rolled out, among other things, a new data system to mine gender data and workshops for staff to share and see good work.

In our experience, conventional evaluations were never this challenging, but neither were they so motivating and useful.

THREE MYTHS BUSTED

To draw more-general lessons, the IDRC worked with an independent specialist to conduct a statistical meta-analysis using blinded data (see ref. 9 for a review). We aggregated results from our 7 independent evaluations of 170 components from 130 discretely funded research projects in natural and social science, undertaken in Africa, Asia, Latin America, the Caribbean and the Middle East[10]. This revealed three things.

Southern-only research is high quality. Research housed wholly in the global south proved scientifically robust, legitimate, important and well-positioned for use. Researchers in the region scored well across each of these criteria (higher, on average, than the northern and north–south-partnered research in our sample). In other words, those most closely linked to a particular problem seem to be well placed to develop a solution. (See Figure S3 in Supplementary Information.)

This finding challenges assumptions that researchers in the north automatically strengthen the capacity of partners in the south[11]. There are many positive reasons to support north–south research partnerships, but the data suggest that we must be strategic to optimize their impact.

Capacity strengthening and excellence go hand in hand. Too many funders assume that research efforts in which teams receive training and skills development inevitably produce poor-quality research. The meta-analysis found no such trade-off. In fact, we found a significant positive correlation between scientific rigour and capacity strengthening.

This suggests that research requiring a focus on capacity strengthening need not be avoided out of a desire for excellence. Indeed, it implies that the two can go hand in hand.

Research can be both rigorous and useful. In the fast-paced world of policy and practice, findings need to get to the right people at the right time, and in ways that they can use (see 'Co-producing climate adaptations in Peru'). We often hear of tension between sample saturation or trial recruitment and the decision-making cycle of policymakers or industry implementers. Happily, the meta-analysis found a strong positive correlation between how rigorous research is and how well it is positioned for use.

This finding builds the case for investing in scientific integrity, in even the most applied and translational programmes.

FOUR CONCERNS

We have four main concerns about RQ+ and how it can be refined and adapted for broader application.

First, bias is baked into our study. We used our own tool to examine research we had already supported. RQ+ focused our post-hoc evaluations on the values that ▶

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



JIM RICHARDSON/NATIONAL GEOGRAPHIC CREATIVE

Farmers in Pampallacta, Peru, inspect harvested potatoes.

CASE STUDY

Co-producing climate adaptations in Peru

More than 500,000 people live in the Mantaro Valley in central Peru, where agriculture is the main source of income. The valley's small-scale farmers provide most of the vegetables and grains consumed in the capital, Lima, but are struggling to respond to the increasing frequency and intensity of extreme droughts, heavy rainfalls and frosts.

Using new and creative combinations of physical measurements and participatory engagement methods such as community mapping, the Geophysical Institute of Peru in Lima is providing a clearer picture of how the climate has changed in the region. This research is informing local policy and guiding adaptation actions. The project mapped hotspots across the region that were susceptible to climate change, and convened discussions with farmers and fishers about how they could adapt schedules and techniques to minimize its impact.

The team did not rush to publish the research in top-tier Western journals, partly because of the English-language barrier but largely because of the urgency of the problem. The research outputs needed to be immediately understandable and usable, so the team rapidly published its findings in working papers and reports (many of which were collected in a Spanish-language book[13,14]). These were immediately accessible to those in local government who needed the evidence to steer the response. As such, predominant metrics do not capture the value of this work.

The RQ+ review shone a different light on this project and its achievements. It scored highly for integrity (including innovative blending of techniques for knowing the climate), for being legitimately grounded in local needs and knowledge, for addressing an urgent problem, and for focusing on uptake and action. **J.L. & R.M.**

matter to our organization. The method examines our objectives and priorities, as we define them. Some would counter that it reifies them.

Second, this tool, much like all others, could have a distorting effect. For instance, by asking reviewers to examine integrity and legitimacy — issues that we identify as fundamental to our success — we turned their attention away from other factors, such as productivity (volume of publications and outputs) and cost-efficiency.

Third, there is the risk that RQ+ results become isolated if they are not comparable with the prevailing measures of research quality used by the global research enterprise. Is RQ+ just another demanding hurdle for researchers in the global south? That's a question we are still working to answer.

Fourth, RQ+ costs more and takes longer than asking two or three peers to offer their opinions. Our hunch is that it takes almost twice as much time and money, largely because it requires empirical data collection by the evaluators. For us, that is time and money well spent: the results help us to hone our approach to funding and engagement.

These concerns will guide our efforts to improve RQ+, as will input from our peers and partners.

MORE LIKE THIS

What next? If the trillions of dollars being invested in research globally each year[12] are to make a difference, we must do better than crude quantification of citations, as the Leiden Manifesto[1] and the San Francisco Declaration on Research Assessment[2] have made clear.

We believe RQ+ presents a practical solution. The approach and findings of our meta-analysis now need replication in other contexts. At IDRC, we are planning another retrospective assessment in 2020. We are excited by what progress and shifts it might uncover. We are already looking at ways we can use RQ+ for grant selection, monitoring the progress of individual projects, and communicating our organizational objectives to funding partners and applicants.

Similarly, we encourage other funders and institutions to improve their evaluations in three ways: consider research in context; accept a multidimensional view of quality; and be systematic and empirical about evidence collection and appraisal. It's time science turned its greatest strengths on itself — experiment, appraise, debate and then improve. ■

Jean Lebel *is president and* **Robert McLean** *is senior programme specialist at the International Development Research Centre, Ottawa, Canada.*
e-mails: jlebel@idrc.ca; rmclean@idrc.ca

1. Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. & Rafols I. *Nature* **520**, 429–431 (2015).
2. Annual Meeting of the American Society for Cell Biology. *San Francisco Declaration on Research Assessment* (ASCB, 2012).
3. Canadian Institutes of Health Research. *Evaluation of CIHR's Knowledge Translation Funding Program* (CIHR, 2013).
4. Wilsdon, J. *et al. The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management* (Higher Education Funding Council of England, 2015).
5. Holmes, B. *The Rise of the Impact Agenda.* Fuse International Conference on Knowledge Exchange in Public Health, Newcastle, UK (27 April 2016); available at https://go.nature.com/2yujm39
6. Amano, T., González-Varo, J. P. & Sutherland, W. J. *PLoS Biol.* **14**, e2000933 (2016).
7. Amano, T. & Sutherland, W. J. *Proc. R. Soc. B* **280**, 20122649 (2013).
8. Ofir, Z., Schwandt, T., Duggan, C. & McLean, R. *Research Quality Plus (RQ+): A Holistic Approach to Evaluating Research* (IDRC, 2016).
9. Gurevitch, J., Koricheva, J., Nakagawa, S. & Stewart, G. *Nature* **555**, 175–182 (2018).
10. McLean, R. & Sen, K. *Making a Difference in the Real World? A Meta-Analysis of Research for Development* (IDRC, 2018).
11. Bradley, M. in *Putting Knowledge to Work* (ed. Mougeot, L. J. A.) Ch. 2 (Practical Action, 2017).
12. '2017 Global R&D Funding Forecast' *R&D Mag.* (Suppl.) Winter 2017, 3–6 (2017).
13. Instituto Geofisico del Peru. *Eventos meteorológicos extremos (sequías, heladas y lluvias intensas) en el valle del Mantaro*. Vol. 1 (IGP, 2012).
14. Instituto Geofisico del Peru. *Manejo de riesgos de desastres ante eventos meteorológicos extremos en el valle del Mantaro.* Vol. 2 (IGP, 2012).

Supplementary information accompanies this article: see go.nature.com/2ja2dfr.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



BEN SKLAR/*THE NEW YORK TIMES*/REDUX/EYEVINE

People from sexual and gender minorities are often under-represented in science.

LGBTQ scientists are still left out

Support from mainstream diversity initiatives would create a virtuous circle of visibility and benefit science for all, urges **Jon Freeman**.

At university in New York City in the 2000s, a professor warned me that I wouldn't get into any PhD programmes if I kept "looking" the way I did. During a single tenure-track job interview in 2011, 13 people asked me: "Do you have a wife?" And when I was an assistant professor, a colleague pulled aside a candidate for a postdoctoral position in my lab to let him know that I'm gay, just in case it would be a problem.

I doubt these people had bad intentions; much has changed since 1975, when gay men and lesbians were still banned from federal employment in the United States.

After all, this week, we mark the first International Day of LGBTQ+ People in Science, Technology, Engineering and Maths (STEM) with plenty of mainstream sponsors, from the American Association for the Advancement of Science to the Wellcome Trust. But heteronormative assumptions can still create less conscious forms of bias, and an unwelcoming environment that puts scientists from sexual and gender minorities (LGBTQ) at a disadvantage.

And science should care more. People who identify as LGBTQ are leaking out of the scientific pipeline in similar ways to women and those from minority ethnic groups. But many initiatives to increase diversity do not support them. Including LGBTQ people in diversity initiatives would foster their representation, and it could bring in perspectives that improve science itself.

LESS VISIBLE LOSSES

Research on LGBTQ people in STEM is scarce, and complicated. Data are hard to collect, not least because sexual identity can be fluid or deliberately concealed. What studies there are, are sobering. Estimates suggest that LGBTQ people are 17–21% less represented in STEM fields than expected[1,2]. Male undergraduates from sexual minorities are much more likely than their straight counterparts to drop out of STEM degrees (see 'Leaky pipeline'), even though they're more likely to pursue practical research experience[3]. In fact, they are dropping out of STEM degrees at a higher rate than women overall[3].

When LGBTQ people continue in STEM, they report more negative workplace experiences than do their counterparts in other industries, or than do non-LGBTQ scientists[1]. Among sexual-minority STEM faculty members who are out at work, 69% report feeling uncomfortable in their department[4].

Of course, the problems run much deeper than science. In the United States, as elsewhere, LGBTQ people face significant disadvantages; these disproportionately affect LGBTQ people of colour. More than one-third of US states lack statutes to protect people from being fired or denied promotion because of their sexual orientation or gender identity (see go.nature.com/2l6akrj). Young LGBTQ people often face rejection or outright abandonment by their parents, and are much more likely to be homeless than are young non-LGBTQ people; LGBTQ adults are more vulnerable to poverty.

Invisibility is a problem for LGBTQ scientists. In my experience, most undergraduates do not know of a single LGBTQ scientist. Nor do they often realize that the founders of entire scientific disciplines, such as the father of artificial intelligence, Alan Turing, were gay or that they faced persecution — for Turing, a sentence of chemical castration, believed to have contributed to his suicide. When I took my first faculty job in 2012, I was, to my knowledge, the only gay assistant professor on campus. As someone starting out, I decided not to actively disclose my sexual orientation to students. Still, rumours spread. Undergraduates in entirely unrelated areas such as engineering or chemistry came to my office under the guise of a vague connection to my field of social neuroscience. Their real motive, left unstated, was to see proof of a gay scientist.

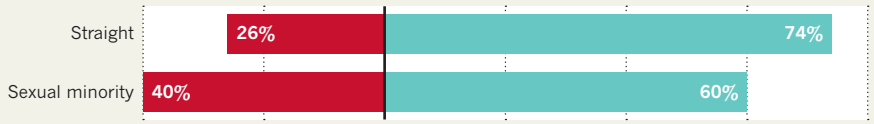
I understand. I didn't meet a senior gay scientist until two years into my faculty career. At a small conference in the mountains, a former mentor advised me to seek out a senior scientist to discuss mutual research interests. At first I was embarrassed that I'd never come across his work, but I soon recognized that we had no research overlap. We didn't even study the same species. As we spoke, I realized what my mentor saw that we had in common. Still, when the scientist casually but deliberately slipped a mention of his husband into conversation, I was shocked. I told him that I had never met a senior gay scientist before, and he said he rarely met any gay scientists at all. We went hiking for the rest of the afternoon. Just being able to talk science with a more senior researcher who was 'like me' was a powerful signal that I had a place in the scientific community.

Common advice to retain and recruit a diverse scientific workforce is to make sure that students, trainees and faculty members from under-represented groups can connect with other members of those groups. But a 2013 survey suggests that more than 40% of LGBTQ workers in STEM are not out to colleagues[5]. The survey also found that ▶

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LEAKY PIPELINE

Men from sexual minorities who start university specializing in science, technology, engineering or maths (STEM) are less likely than straight men to still be in STEM-focused degrees after four years.



*Sample size for straight men was 1,619; for men from sexual minorities it was 124.

SOURCE: B. HUGHES

most respondents could not name a single LGBTQ faculty member at the universities where they got their degrees. The culture in STEM fields might be at fault; so might university policies. LGBTQ scientists are less likely to be open at institutions that do not offer same-sex partner benefits or support name changes during a gender transition[5].

In science, where our personal lives already take a back seat, it can feel unprofessional or career-damaging to be open about something as personal as our LGBTQ identity, and no scientist should feel pressure to do so. But without visibility, other scientists will not benefit from a sense of belonging and inclusion.

Now that I have tenure, I feel more comfortable. Yet I struggle to strike a balance between openness and professionalism. A casual reference to 'my husband' would suffice, at least in regions where it is safe to do so, but I'm single. Displaying a rainbow symbol in my office could work, but I worry that some will think I'm calling for attention or being 'political'. And I don't want my sexual orientation to be my defining characteristic, either; I want it to be incidental. These days, I'll usually invoke a pop-culture reference that lets others infer that I'm gay. Most scientists — LGBTQ or otherwise — can take the hint.

I'm heartened by Twitter and other social-media platforms that allow LGBTQ scientists to find each other. It's inspiring to see those promoting visibility by submitting their bios and stories to sites such as 500 Queer Scientists. However, I worry that this visibility might give those early in their careers a misleading picture. Science can be slow to change. In many STEM departments, all tenured faculty members vote on each tenure-track recruit, someone they expect to be a colleague for the rest of their working lives. In a market that has far too few opportunities as it is, LGBTQ scientists suffer even more when job-search committees and voting faculty members assume that LGBTQ scientists won't 'mesh' with their department or settle outside urban areas in the long term.

LGBTQ trainees should not have to worry about making the mistake of their career because of lazy heteronormative logic. As a job candidate, each of the 13 times a faculty member asked me whether I had a wife — technically in violation of university policy — I had to make a choice. I could politely correct their assumption, probably embarrassing them and damaging our rapport. Or I could answer 'no,' deflect any other personal questions, feel dishonest and forgo a stronger personal connection with someone who might be hiring me into a position that could last into my retirement. Either way, there is the potential for negative impact.

CALL TO ACTION

Advocacy groups such as Out in STEM and the US National Organization of Gay and Lesbian Scientists and Technical Professionals are doing important work to connect LGBTQ scientists. However, the most natural points of intersection occur in specific fields — where physicists meet other physicists and biologists meet other biologists. Scientific societies and conferences need to take on this role more proactively. So do research institutions and funding agencies.

Diversity programmes to develop the scientific workforce at both the US National Science Foundation (NSF) and the US National Institutes of Health (NIH) consistently leave out LGBTQ people. Although the NSF's review criteria for trainee fellowships include the "development of a diverse, globally competitive STEM workforce" (go.nature.com/2micyjf) and the NIH's diversity definition is predicated on evidence of under-representation (go.nature.com/2k6shpk), neither agency makes specific mention of LGBTQ people. The NIH explicitly accounts for minority ethnic groups, people with disabilities, women and people from low-income families, but not LGBTQ people. In fact, the NSF's analyses of STEM participation, widely used by funding agencies and universities, do not even track LGBTQ people[6]. Thus, the data that could inform policy are not being collected.

"LGBTQ trainees should not have to worry about making the mistake of their career because of lazy heteronormative logic."

Similarly, although most universities have non-discrimination policies for LGBTQ people, their diversity initiatives for recruiting faculty members and trainees typically omit us. Increasingly, faculty postings ask applicants for diversity statements, and trainees often wonder whether they should disclose their LGBTQ identity. The problem is that search committees might make assumptions that hurt LGBTQ applicants. And even if some members of search committees believe LGBTQ people provide a valuable form of diversity, that matters little if those candidates won't receive benefits from the university's initiatives.

Including those who identify as LGBTQ in mainstream diversity initiatives would encourage LGBTQ representation and send an official signal to the scientific community. Those appointing faculty members and lab personnel, or reviewing graduate applications, might be more willing to give us the benefit of the doubt (or, better, question the basis for their doubts). All of this, in turn, would lead to more peers and role models and ease challenges of invisibility.

I don't think including LGBTQ people in diversity efforts lessens the importance of these initiatives for women and people from minority ethnic groups; these are real and urgent. Indeed, broadening diversity efforts could have synergistic impacts. For instance, LGBTQ people are more likely to be out in STEM fields that have greater representation of women[5].

And improving LGBTQ representation could help scientific research itself. Many studies have shown[7] that gender, racial, geographic and political diversity, as well as diverse personal experiences, all bring unique perspectives that improve group decision-making, company performance and the quality of scientific work.

If we are to eliminate the insidious ways in which bias, even in well-intentioned people, continues to hinder our career attainment, we must recognize and avoid it in our community and our policies. An LGBTQ graduate student I met recently relayed on Twitter his experience of asking a full room of scientists why there are so few out faculty members. They argued that there is no point in being out in science: acceptance is implicit and whether someone is LGBTQ is irrelevant.

The data — and my experience — show otherwise. ■

Jon Freeman *is associate professor of psychology and neural science at New York University in New York City and director of the Social Cognitive & Neural Sciences Lab. e-mail: jon.freeman@nyu.edu*

1. Cech, E. A. & Pham, M. V. *Soc. Sci.* **6**, 12 (2017).
2. Cech, E. *Proc. 2015 Am. Soc. Eng. Educ. Natl Conf.* https://doi.org/10.18260/p.24431 (2015).
3. Hughes, B. E. *Science Adv.* **4**, eaao6373 (2018).
4. Patridge, E. V., Barthelemy, R. S. & Rankin, S. R. *J. Women Minorities Sci. Eng.* **20**, 75–98 (2014).
5. Yoder, J. B. & Mattheis, A. *J. Homosexuality* **63**, 1–27 (2016).
6. National Science Foundation. *Women, Minorities, and Persons with Disabilities in Science and Engineering* (NSF, 2017); available at http://go.nature.com/2yrkdq5
7. Galinsky, A. D. *et al. Perspect. Psychol. Sci.* **10**, 742–748 (2015).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

UNESCO



Workers move part of a statue during the relocation of the Abu Simbel temples in Egypt in the 1960s.

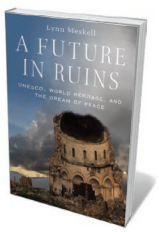
HERITAGE SCIENCE

Revisiting the culture wars

Andrew Robinson admires an exploration of UNESCO's highs and lows.

In 1945, three months after the end of the Second World War, delegates from 44 countries gathered in London. Britain's prime minister, Clement Attlee, told the conference that "the peoples of the world are islands shouting at each other over seas of misunderstanding". The delegates proposed another way: the United Nations Educational, Scientific and Cultural Organization (UNESCO). The agency, based in Paris, was dedicated to preserving peace by promoting the international exchange of ideas. Its first director-general was the British biologist, humanist and amateur archaeologist Julian Huxley.

In *A Future in Ruins*, archaeologist Lynn Meskell offers an institutional ethnography of UNESCO. The organization's broad remit ranges from publishing to promoting women in science, but Meskell focuses exclusively on its role in protecting world heritage and archaeology, particularly through the 1972 World Heritage Convention. Inevitably, this role has been highly political. UNESCO's mission was "to end global conflict and help the world rebuild materially and morally", Meskell observes. Yet increasingly, she argues, its efforts are caught up in the proliferation and prolongation of local conflicts and tensions.



A Future in Ruins: UNESCO, World Heritage, and the Dream of Peace
LYNN MESKELL
Oxford University Press (2018)

Witness the World Heritage sites at, for instance, Angkor in Cambodia; Bodh Gaya in India; Kandy in Sri Lanka; Palmyra in Syria; and Timbuktu in Mali. The 1992 inscription of Angkor in the World Heritage List was supported by exiled sympathizers of the genocidal Khmer Rouge regime, hoping to bolster territorial claims. As nations jostle on and around UNESCO's committee to achieve listing, it is clear that designation is about more than past glory. It is also a way of ensuring future rewards, most obviously the boosting of tourism.

The daunting nature of the agency's goal was obvious from the start. Huxley was under no illusions about it, referring to the "impossibility of UNESCO producing the rabbit of political peace out of a cultural and scientific hat". In 1948, Huxley was eased out of office at the behest of the US delegation (possibly owing to his left-wing humanism). Classicist Gilbert Murray subsequently predicted that UNESCO was destined to be "a confused mixture of success and failure". He had a unique perspective. From 1922 to 1939, he had served on the agency's predecessor, the League of Nations' dispute-ridden International Committee on Intellectual Cooperation, which included scientific luminaries ▶

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



PAUL CHESLEY/NGC/GETTY

Buddhist monks in the Angkor Wat temple, Cambodia, which has been a World Heritage site since 1992.

Marie Curie, Albert Einstein and Hendrik Lorentz. As Meskell concludes, Murray's prediction rings true seven decades on.

Nonetheless, UNESCO's successes have been impressive. It came to the rescue of Venice in 1966, when devastating flooding threatened swathes of the Italian city's cultural treasures, valued at US$6 billion. Under the agency's auspices, Japan funded and directed the restoration of Indonesia's Buddhist temple of Borobudur. Its greatest triumph is probably still its much-publicized 20-year Nubian campaign. Launched in 1959, this aimed to rescue ancient Egyptian and Sudanese sites from flooding, caused by the 1960–70 construction of the Aswan High Dam on the Nile. Meskell devotes an informative chapter to this extraordinary feat. Some 23 temples, tombs, early Christian churches and rock-hewn chapels were dismantled and relocated — most famously the temple of Rameses the Great at Abu Simbel, at a cost of $70 million.

Even here, serious internal problems surfaced. The British government refused to contribute because of the 1956 Suez crisis, when it came to blows with Egypt over control of the eponymous canal. However, many British Egyptologists and archaeologists, such as Mortimer Wheeler, lent support. UNESCO, preoccupied with rescuing known sites scheduled for flooding, did little to promote new archaeological investigation; instead, the work was independently financed by some 40 expeditions, and poorly coordinated as a result. Meanwhile, Nubians living in flooded areas were moved far from their ancestral homes, traditional means of subsistence and ways of life.

In the years that followed, UNESCO's heritage work was repeatedly weakened by political pressures from member nations, a focus on monumental recovery rather than archaeological discovery, and a lack of participation by indigenous people. An example is Hampi in India — the remains of Vijayanagara, the capital of the last great Hindu kingdom — which was declared a World Heritage site in 1986. In 2011, the Indian government, backed by the Archaeological Survey of India, cleared modern infrastructure from the site, evicting residents and blocking UNESCO monitoring.

Meskell offers a trenchant critique of how UNESCO's aim of preventing war sits oddly with projects commemorating sites associated with violence. In 1978, the Senegalese island of Gorée became a World Heritage site marking human exploitation in the international slave trade. In 1979, the former Nazi concentration camp at Auschwitz–Birkenau was listed. In 1996, the Hiroshima Peace Memorial — the only structure left standing after the atomic bombing of 1945 — was added. And in 2016, the medieval Armenian city of Ani in Turkey became a World Heritage site despite the 1915–22 Armenian genocide, which Turkey has denied. Now, Thailand is seeking UNESCO endorsement for the Burma–Siam Railway (known as the Death Railway); the country promotes it as a Second World War tourist attraction even though it was constructed by Japan using forced labour and prisoners of war.

Meskell notes that international recognition enshrines only one version of history. That, she argues, renders UNESCO complicit in recontextualizing "episodes of illegal occupation, atrocities, war crimes, and even genocide, while the victims are left to relive the trauma. This is the dark side of heritage branding." Moreover, as tourism burgeons and the World Heritage brand has increased in value to member countries and listed sites, UNESCO's overall budget of $250 million has been increasingly squeezed. It will probably fall further if the United States withdraws its membership for the second time, at the end of 2018. All this undermines the agency's conservation and management of sites and its educational and scientific mission.

UNESCO's workforce consists increasingly of volunteers. In 2017, some 600 unpaid interns from all over the globe thronged its headquarters, among consultants and a reduced staff trying to manage two or three posts per person. It is a sad decline from the aspirations of that post-war moment in 1945. And given the increasingly aggressive tone of international diplomacy, it is also worrying. ■

Andrew Robinson *is the author of seven books on archaeology, including* Cracking the Egyptian Code *and* The Indus: Lost Civilizations.
e-mail: andrew@andrew-robinson.org

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

EVOLUTION

Travels on the backbone

Chris Lowe lauds a study of vertebrate origins that brings us up to date with a shifting field.

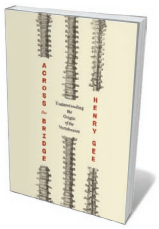
Some of the great remaining mysteries in zoology concern origins — of multicellularity, complex nervous systems, life cycles and sex, for example. The evolutionary origin of vertebrates is among the most intractable of these, despite more than a century of work spanning a range of disciplines and animal groups.

In *Across the Bridge*, Henry Gee reviews the most recent research in this area. Gee (the senior editor responsible for palaeontology and evolutionary development at *Nature*) synthesizes contributions from anatomy, developmental biology, genomics, palaeontology and the study of evolutionary relationships using DNA-sequence data. He also puts forward his own ideas on this fascinating conundrum.

The book is a follow-up to Gee's 1996 *Before the Backbone*, which was one of the reasons I decided to pursue research into the origins of chordates (the group including vertebrates, the fishlike lancelets and marine invertebrates called tunicates). At the time, the discipline was intimidating, involving impenetrable papers and a dizzying array of contradictory hypotheses. Gee tamed and synthesized the literature to lay out the history and logic of the most significant hypotheses, such as Walter Garstang's 1894 auricularian theory. (This posited that the chordate body plan evolved through transformation at the larval, rather than adult, stage.) *Before the Backbone* inspired novices like me to get excited about the challenges in vertebrate origins.

Is a major new summary of the field warranted, just two decades on? In my view, yes. (Among many others in the field, I encouraged Gee to write this book, and he thanks me for so doing in the preface.) There has been significant progress across disciplines on research into important, yet previously neglected, animal groups at crucial positions in the evolutionary tree. Along with tunicates and lancelets, these include hagfish, lampreys and marine worms called hemichordates. The publication of the genomes of two hemichordates — the acorn worms *Saccoglossus kowalevskii* and *Ptychodera flava* (O. Simakov *et al. Nature* **527**, 459–465; 2015) — rounded out the representative genomes of almost all major animal groups with direct relevance to vertebrate origins. That draws a line under a 20-year burst of productive research.

Some of the problems that once stalled progress — such as uncertainty over which groups belong in the deuterostome lineage of animals alongside chordates — have largely been resolved. Others are as intractable as ever, including where to place key fossil groups such as the curious vetulicolians, which lived during the Cambrian period, some 541 million to 485 million years ago. Nevertheless, Gee effectively develops his own evolutionary scenario. He traces vertebrate origins from the common ancestor of chordates, echinoderms (which include starfish and sea urchins) and hemichordates, deep in animal history, to establish a starting point from which innovations of the first chordates arose. Having sketched that picture, he traces the evolution of definitive vertebrate characteristics by mining rich new studies from neglected chordate groups, from lancelets and tunicates to hagfish and lampreys. His well-argued rationale draws from anatomy, palaeontology and molecular genetic data.



Across the Bridge: Understanding the Origin of the Vertebrates
HENRY GEE
University of Chicago Press (2018)

The book is not a revision of *Before the Backbone*. But one of its goals seems to be to examine how the most influential classical hypotheses fare in the light of new data, which is one of Gee's unique perspectives. He begins with a group-by-group examination, from echinoderms to hagfish and a few non-deuterostomes. In the most effective section, he strips vertebrates down to their parts, such as the nerve cord, notochord (forerunner of the vertebral column) and neural crest (a group of embryonic cells). He even delves into the largely ignored gut and viscera.



NOAA OKEANOS EXPLORER PROGRAM

Hagfish have been neglected in vertebrate studies.

Gee discusses how data from living and fossilized hemichordates and echinoderms have facilitated the search for the origins of the defining chordate anatomies. He highlights how palaeontological, developmental and genomic data now all support the idea that the common ancestor of chordates, hemichordates and echinoderms had pharyngeal gill slits for filter feeding. That gives us a glimpse of the early chordate ancestor, which lived around 600 million years ago. Other features, such as a complex brain, probably emerged much later. Having established a hazy picture of the earliest chordates, Gee focuses on building vertebrates and their defining features from the basic chordate body plan, for example through spectacular innovations in the vertebrate head.

Gee considers the evolution of each characteristic independently and as a component of the vertebrate body plan. He includes alternative interpretations and areas in which data are weak or missing, allowing the reader to think hard about some of the more speculative parts of his arguments. Various bizarre animals illustrate particular points. For example, the strange gelatinous marine invertebrates called larvaceans secrete a mucus 'house' for filter feeding, which they shed and rebuild every few hours. They are an extreme example of how our closest invertebrate relatives have taken their own evolutionary tangent, partly driven by genome simplification.

My main criticism of *Across the Bridge* is that for a subject this visual, I would have liked more illustrations. It is, after all, the oddness of the main groups and their stem fossils that has made vertebrate origins such a difficult nut to crack.

Do I buy Gee's original synthesis? Mostly. The morphological divide between chordates and their closest relatives remains perilously large, and even with new advances, Gee's hypothesis is a good contender; but it is not the only one, so the bridge to early chordates remains a little wobbly. (An alternative treatment of recent progress is outlined in Noriyuki Satoh's 2016 *Chordate Origins and Evolution*.) But *Across the Bridge* is a deft and well-argued distillation of how advances have shifted the field to the point of dispatching some of the most influential and elegant classical hypotheses, and it is a bold attempt at developing a new synthesis. It thereby deepens understanding of our own evolutionary origins. ■

Chris Lowe *is an associate professor at Hopkins Marine Station in the Department of Biology at Stanford University, California. e-mail: clowe@stanford.edu*

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Correspondence

Equalize access to medical advances

We contend that governments, intergovernmental agencies and non-governmental organizations should ensure that all patients have access to 'orphan' drugs for rare diseases and to personalized medicine services, not just those who can afford to pay.

Within the United States and the European Union, tax incentives accrue to companies that are engaged in orphan-drug development. Start-ups and large multinational companies, which tend to seek enhanced returns for investors, could soon emerge as specialist providers of medicines to those who can afford personalized health-care options. Less-affluent patients would then be at risk of being excluded from advances in molecular diagnostics and in genetic-modification strategies for disease avoidance.

Technologies often outpace preparations for their application in health care, including in such fields as regenerative medicine (P. Marks and S. Gottlieb *N. Engl. J. Med.* **378**, 954–959; 2018). In this instance, too, we might have only a narrow window of opportunity to influence international cultures, regulations and investment to ensure equitable benefit for all.

Colum P. Dunne, **Suzanne S. Dunne** *Graduate Entry Medical School and Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick, Ireland.*
colum.dunne@ul.ie

Italy squeezes out wet biology research

Italy's wet biology research is in sharp decline, thanks to scarce funding and stifling bureaucracy. These forces are driving young scientists away from the laboratory bench and out of academia.

Basic experimental-biology research has been viewed as a bad bet by Italian funding agencies over the past couple of decades. The situation is made worse by anti-corruption rules introduced last year by the government. These further complicate the already cumbersome administration procedures of public institutions. It can take weeks or even months to acquire a chemical reagent or a piece of equipment.

To boost their research credentials and improve their career opportunities, Italian biologists are therefore abandoning mechanistic bench research in favour of *in silico* investigations. These generate data rapidly and do not require expensive laboratory equipment or compliance with a tangle of safety regulations.

We need to reverse the situation so that future Italian governments want to invest in basic research. Although it is encouraging that the former Italian government earmarked a generous sum of €391 million (US$456 million) for its 2017 3-year Research Projects of Relevant National Interest, this is currently a one-off investment that could be discontinued by future governments.

Expanding research-evaluation criteria from citation counting would help — for example, by including the intellectual property of a scientific work, its actual impact within a specific scientific community and the novelty of experimental approaches that maximize meaningful output.

Davide Zannoni *University of Bologna Alma Mater Studiorum, Italy.*
davide.zannoni@unibo.it

University fought a researcher's corner

A university's leaders might disown researchers in the face of allegations of unethical practice out of fear for the establishment's reputation (see, for example, *Nature* **558**, 13–14; 2018). Three years ago, I had the opposite experience.

I was accused of animal maltreatment on an Italian national television programme, on the basis of images that had been taken illegally by an undercover activist. An animal-welfare organization brought charges against me to a tribunal in Rome, claiming that I had violated Italy's animal-protection law.

My university press office meticulously investigated all of my lab's papers and protocols with the university's official veterinary surgeon and wrote to the director of the TV programme, emphasizing that all procedures in the lab had conformed with the law. The press office also relayed a letter from me to all senior academic staff that set out my version of events.

Thanks to this support, I received an unexpectedly positive international response. The university formally conducted my legal defence at the tribunal. I have since received authorization from the Italian Ministry of Health to continue with my neuroscience research on non-human primates.

Reciprocal communication between researchers and their institutional leaders is crucial for resolving such disputes. My case illustrates how fair decisions can be reached by institutions in critical situations after careful assessment of the facts, unprejudiced by their potential impact on an institution's standing.

Roberto Caminiti *Sapienza University of Rome, Italy.*
roberto.caminiti@uniroma1.it

Evaluation woes: DORA responds

We understand John Tregoning's frustration with the slow progress in eliminating undue reliance on journal impact factors in research assessment (*Nature* **558**, 345; 2018). However, his claim that no alternative has emerged overlooks the many examples of good practice implemented by funders, learned societies, research institutes and individual scientists. The San Francisco Declaration on Research Assessment (DORA) showcases some of these (see https://sfdora.org).

There is no quick fix to the problems arising from the misapplication of impact factors in evaluating researchers. Assessing research across disciplines and geography is complex, and DORA has been careful to call out the need to address that complexity. Settling for the least-bad option is not the answer.

We agree with Tregoning's suggestions for diversifying measures of academic success. Examples include data and methods sharing, teaching, collegiality and public engagement. These are in line with initiatives that are under way. Universal agreement on best practices has yet to emerge, but progress is being made.

We appreciate the stresses of shaping a career in a highly competitive system, especially when the rules of the game are changing. But that change is necessary and will take time. DORA is committed to speeding things along so that we can provide the clarity Tregoning calls for. See also page 23

Anna Hatch *DORA, Bethesda, Maryland, USA.*
Stephen Curry* *DORA and Department of Life Sciences, Imperial College London, UK.*
ahatch@ascb.org
**S.C. declares competing non-financial interests; see go.nature.com/2tbl3us for details.*

CONTRIBUTIONS

Correspondence may be submitted to **correspondence@nature.com** after consulting the author guidelines and section policies at **http://go.nature.com/cmchno**.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



JIM BRANDENBURG/MINDEN PICTURES/FLPA

Figure 1 | A glacier at Mount Robson Provincial Park, British Columbia, Canada. An analysis by Schildgen and colleagues[3] confirms that the rate of mountain erosion by glaciers has increased during the past few million years in certain places (such as in British Columbia) in response to climate cooling, but casts doubt on the idea that this was a global effect.

EARTH SCIENCE

Global erosion by glaciers revisited

Mountain erosion is thought to have sped up globally over the past few million years as the climate cooled and glaciers grew. A reassessment of the data suggests that this acceleration was limited to just a few regions. SEE LETTER P.89

ERIC KIRBY

A long-standing question in Earth sciences has been whether the climatic cooling that has occurred over the past 4 million to 6 million years — and the consequent growth of glaciers in temperate mountain ranges — accelerated the global rate at which Earth's surface has eroded. The answer has important implications for our understanding of the interplay between climate, landscape topography and plate tectonics[1]. In a 2013 paper in *Nature*, Herman *et al.*[2] evaluated the thermal histories of thousands of rocks sampled from dozens of mountain ranges, and concluded that rapid cooling of these samples heralded an acceleration in erosion driven by mountain glaciers. On page 89, Schildgen and colleagues[3] report that the apparent changes in erosion rates over time calculated in that study were largely the result of data being combined from samples that had separate erosion histories and that took different times to reach Earth's surface.

Mountainous landscapes are the manifestation of the competition between forces that arise from the collision of tectonic plates, which thicken the crust and build high-altitude topographical features, and the destructive action of erosion by wind, water and ice, which transforms bedrock into sediment and transports rock detritus into sedimentary and ocean basins. The rate at which erosion occurs globally is a key factor in the climate system over geological timescales. Volcanic activity supplies the atmosphere with carbon dioxide, but the chemical weathering of silicate minerals consumes CO_2; the balance between these two processes affects atmospheric CO_2 levels, which in turn regulate temperatures over millions of years[4]. If the supply of silicates at the surface grows, an increase in weathering might effectively scrub the atmosphere of CO_2, leading to a decrease in global temperatures[5].

Earth's climate began to cool about 15 million years ago[6], a trend that intensified approximately 6 million years ago and led to the expansion of both continental ice sheets and mountain glaciers. Apparent increases in the global accumulation of sediment in oceanic and continental basins during the same period[7–9] has often been cited as evidence that sediment production, sediment transport and the erosive action of glaciers became more effective as the climate cooled.

However, stratigraphic records also reveal episodes when no sedimentary deposition

© 2018 Springer Nature Limited. All rights reserved.

occurred. These hiatuses impart a statistical bias that causes sediment accumulation rates to seem greater when calculated from measurements taken at shorter time intervals[10]. Indeed, geochemical studies suggest that the delivery of continent-derived sediment to the global ocean has essentially remained constant over the past approximately 10 million years[11].

This statistical pitfall can be overcome by directly assessing the thermal history of rocks. Rocks in the crust cool as overlying material is removed by erosion at the surface (a process referred to as exhumation), because they are brought nearer to the surface. Several minerals in bedrock contain trace amounts of radioactive isotopes, whose decay generates daughter products. The ability of minerals to retain these products in their crystal lattices depends on temperature, and so the amount of a daughter product that has accumulated in a mineral can be used to work out the time passed since the mineral was at a specific temperature (the blocking temperature). This forms the basis of a technique known as thermochronology. By analysing mineral systems that have different retentivities, the rate at which rock cooled can be estimated.

Herman and colleagues took this idea a step further by developing a method that works out the temperature and exhumation histories that best fit thermochronologic data obtained for multiple rock samples collected within close proximity. This approach assumes that nearby samples have similar exhumation histories, and thus that any differences in thermochronologic age between mineral systems with different blocking temperatures are explained by the systems having different cooling rates over time. For the past five years, this work has stood as the most compelling evidence that the cooling of global climates drove synchronous increases in erosion rate as a direct consequence of the growth of mountain glaciers.

Schildgen *et al.* now demonstrate that the assumption that nearby samples have similar exhumation histories is often not appropriate. In nearly all of the regions that were thought to exhibit an increased exhumation rate over time[2], samples that have been recently exhumed are juxtaposed with older samples at geological boundaries. In the Western Alps, for instance, an abrupt difference in thermochronometer age for samples collected across a fault system reflects the different erosional histories for rocks on either side of the fault. Yet when these data are combined using Herman and colleagues' data-analysis method, the age difference is transformed into an apparent increase in average erosion rate through time for the entire region. Perhaps even more convincingly, Schildgen and colleagues used artificial data sets constructed to represent known geological scenarios to demonstrate that spurious results can occur when data are combined from samples subjected to erosion that was constant in time, but non-uniform in space.

The authors evaluated the other 29 examples of apparent accelerated erosion presented by Herman *et al.*, and found that nearly three-quarters could have resulted from the combination of data from samples that have different erosion histories. They argue that a few of the remaining cases reflect changes in tectonic boundary conditions (changes in fault activity). Only three of the original study sites seem to have experienced climatically driven acceleration in cooling, and each of these was associated with the formation of deep valleys by glaciers (Fig. 1). This re-evaluation suggests that, although glaciers can be an effective agent of landscape erosion, the thermochronology methods used by Herman *et al.* don't have the resolution required to assess whether global increases in glacial erosion occurred during the past few million years.

The finding that some sites do seem to reflect enhanced erosion by glacial ice, however, suggests that the story is not yet complete. As the precision and sensitivity to lower temperatures (shallower depths in Earth's crust) of thermochronology continue to improve, the method might be able to constrain estimates of changes in erosion over relatively short timescales[12]. Similar improvements in the precision with which the age of coarse sediments derived from glacial erosion can be determined[13] might also help to resolve questions of how rapidly such deposits accumulated.

Schildgen and colleagues' analysis reminds us that deformation of Earth's rigid outer shell matters: it deforms in a brittle way, breaking along geological faults that, in turn, produce boundaries between regions that have different rates of uplift. The resulting mountain ranges are subject to localized differences in exhumation. The authors' findings also teach us a broader lesson: as we use increasingly sophisticated analyses of 'big data' to gain insight into global trends in geology, we must not lose sight of the physical processes that operate locally. Incorporating site-specific geological constraints into analyses of global data sets will be essential in future studies.

Knowledge of how Earth responded to past climate change continues to be relevant to our future. Schildgen and colleagues' work removes one of the remaining pillars that supported the hypothesis that cooling climates were responsible for a global acceleration in erosion. The idea that erosion rates over the past few million years might have been less sensitive to cooling than was thought challenges us to re-examine our long-held notions of how erosion and climate systems interact. ■

Eric Kirby *is in the College of Earth, Ocean and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97330, USA.*
e-mail: eric.kirby@oregonstate.edu

1. Raymo, M. E. & Ruddiman, W. F. *Nature* **359**, 117–122 (1992).
2. Herman, F. *et al. Nature* **504**, 423–426 (2013).
3. Schildgen, T. F., van der Beek, P. A., Sinclair, H. D. & Thiede, R. C. *Nature* **559**, 89–93 (2018).
4. Berner, R. A., Lasaga, A. C. & Garrels, R. M. *Am. J. Sci.* **283**, 641–683 (1983).
5. Raymo, M. E., Ruddiman, W. F. & Froelich, P. N. *Geology* **16**, 649–653 (1988).
6. Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. *Science* **292**, 686–693 (2001).
7. Hay, W. W., Sloan, J. L. II & Wold, C. N. *J. Geophys. Res.* **93**, 14933–14940 (1988).
8. Peizhen, Z., Molnar, P. & Downs, W. R. *Nature* **410**, 891–897 (2001).
9. Willett, S. D. *Annu. Rev. Earth Planet. Sci.* **38**, 411–437 (2010).
10. Schumer, R. & Jerolmack, D. J. *J. Geophys. Res. Earth Surf.* **114**, F00A06 (2009).
11. Willenbring, J. K. & von Blanckenburg, F. *Nature* **465**, 211–214 (2010).
12. King, G. E., Herman, F. & Guralnik, B. *Science* **353**, 800–804 (2016).
13. Balco, G. & Shuster, D. L. *Earth Planet. Sci. Lett.* **286**, 570–575 (2009).

CELL BIOLOGY

Actin proteins assemble to protect the genome

The assembly of polymerized actin with motor proteins at DNA breaks in the nucleus supports the mobility and repair of DNA. This finding reveals a layer of regulation that helps to preserve genome integrity. SEE ARTICLES P.54 & P.61

VASSILIS ROUKOS

The protein actin polymerizes to produce filaments that form crosslinked networks in the cytoplasm of cells. These networks support many fundamental biological processes — such as cell movement and division and the intracellular trafficking of molecules. Reports that actin also has functions in the cell nucleus remain controversial[1,2], partly because of the challenges of performing experiments that exclusively perturb the nuclear actin pool without also perturbing actin in the cytoplasm. Two studies[3,4] in this issue now provide the most compelling evidence so far that polymerized actin has roles in cell nuclei in which DNA has been damaged, and that it could be essential for maintaining genome stability.

In cell nuclei, DNA is packaged with

© 2018 Springer Nature Limited. All rights reserved.

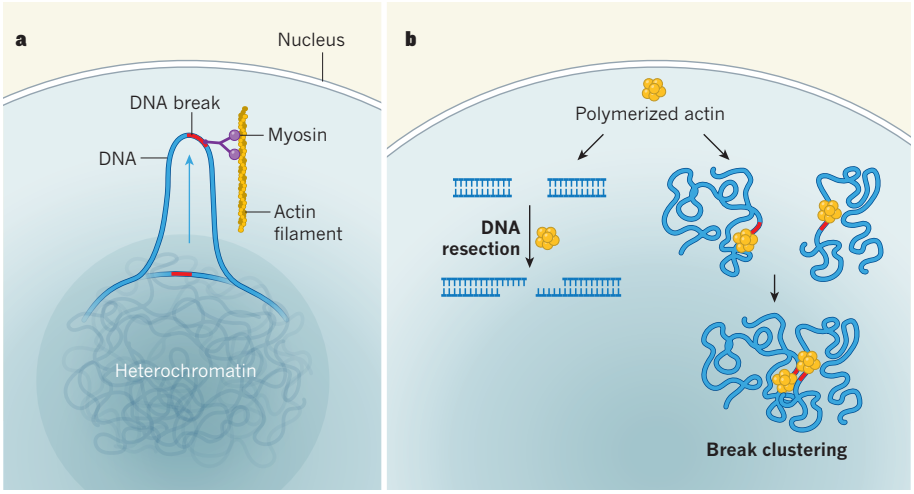


Figure 1 | Nuclear actin polymerizes to preserve genome stability. DNA is packaged with proteins in the nucleus to form a material called chromatin, which is compartmentalized into different domains, including heterochromatin. **a**, Caridi *et al.*[3] report that, in cells of the fruit fly *Drosophila melanogaster*, DNA breaks in heterochromatin are moved to the nuclear periphery to ensure correct repair — motor proteins called myosins 'walk' the DNA breaks along filaments made from polymerized actin protein. **b**, Schrank *et al.*[4] report that, in human cells, polymerized actin promotes the processing (resection) of broken DNA ends in chromatin (the compartments involved were not determined) and facilitates a DNA-repair pathway called homologous recombination (HR, not shown), although the underlying mechanism is unclear. Polymerization of actin also increases the mobility of DNA breaks that will be repaired by HR and the ability of a subset of DNA breaks to form clusters in the nucleus. Prevention of any of the processes shown in **a** and **b** leads to repair defects and increased genomic instability.

proteins to form a material called chromatin, which is subdivided into different domains. One of these domains, known as heterochromatin, is a tightly packed form of DNA characterized by large segments of repeated DNA sequences. These are highly prone to an aberrant form of genome shuffling called ectopic recombination, and therefore present a serious threat to genomic integrity. In fruit flies and mice, heterochromatic DNA that contains double-strand breaks (DSBs) is moved outside this domain to prevent such aberrant recombination[5–7], but the mechanism involved has not been known. Actin filaments form in the nuclei of mammalian cells in response to DNA damage[8,9], but their function in DNA repair has also been unclear.

Caridi *et al.*[3] (page 54) show that a subset of DSBs in fruit-fly cells move from heterochromatin towards the nuclear periphery in long-lasting, directed motions, rather than along random paths. Such directed motion has previously been observed only for a handful of cases in the nuclei of mammalian cells[10–12]. Crucially, the authors provide insight into how this relocation occurs. They find that nuclear actin polymerizes to form filaments at heterochromatic DSB-repair sites, in a process that requires the presence of the protein complex Arp2/3 and its activators (the Scar and Wash proteins). They observe that Arp2/3 promotes the formation of actin filaments that grow from heterochromatic DSBs towards the nuclear periphery. The nuclear motor proteins myosin I and myosin V then 'walk' the repair sites along the actin filaments (Fig. 1a). This walk is triggered by the myosin-activating protein Unc45 after filament formation, indicating that relocation is regulated both in space and in time.

An important finding of Caridi and colleagues is that recruitment of Arp2/3 and myosins at heterochromatic DSBs requires the DSB-repair protein Mre11 and a heterochromatin component, the HP1α protein. This suggests that DSB detection or processing is required for the recruitment of the motor proteins to the damaged sites and reveals why only heterochromatic DSBs are relocated to the nuclear periphery. Notably, the authors demonstrate that, when nuclear actin and myosins are defective in both fruit-fly and mouse cells, the integrity of the genome in heterochromatin is impaired and cells become less able to survive DNA damage, suggesting that actin and myosin have key roles in maintaining genome stability in the nucleus.

DSB repair can proceed by two distinct mechanisms: the error-prone non-homologous end-joining pathway (NHEJ) and the usually error-free homologous recombination (HR) pathway. During NHEJ, broken ends of DNA are simply re-stitched together, whereas HR requires that DNA ends are first processed to generate stretches of single-stranded DNA, which then search for appropriate (homologous) DNA sequences to use as a template for repair. Schrank *et al.*[4] (page 61) report that polymerization of nuclear actin in human cells is specifically required for the efficient repair of DSBs by HR.

Using a cellular system that controllably induces DSBs mainly within euchromatin[13], the authors find that — as in heterochromatic breaks — Arp2/3 is recruited to the damaged genomic sites. However, the Arp2/3 activators recruited to these breaks are different from the ones shown by Caridi *et al.* to be recruited at heterochromatic breaks. This raises the possibility that different regulatory mechanisms promote actin-filament formation in different genomic and chromatin contexts.

Intriguingly, Schrank and colleagues find that Arp2/3 is recruited only at DSBs undergoing HR, whereas the Wiskott–Aldrich syndrome protein (WASP, an Arp2/3 activator) is recruited at breaks repaired by both HR and NHEJ. This observation indicates the existence of a currently unclear regulation mechanism that promotes actin polymerization so that it occurs specifically at breaks undergoing HR. The researchers also show that inhibition of nuclear-actin polymerization leads to a decrease in the processing of DNA ends and in the efficiency of HR repair, placing nuclear actin at the core of the cellular response to DNA damage (Fig. 1b).

Schrank and co-workers' efforts might turn out to be relevant to disease. Wiskott–Aldrich syndrome (WAS) is a disorder characterized by severe immunodeficiency and predisposition to cancers caused by mutations in the *WAS* gene[14] (which encodes WASP). How *WAS* mutations cause malignancies is largely unknown at the molecular level. Schrank *et al.* report that inhibition of WASP leads to defects in HR repair, and that immune cells (lymphocytes) from people with WAS show evidence of defects in DNA-end processing and have higher sensitivity to DNA-damaging agents than do lymphocytes from healthy people. Future studies should investigate whether the repair defects associated with WASP malfunction in the nucleus contribute to the malignancies in people with WAS.

The authors further show that perturbations of actin polymerization in the nucleus lead to a decrease in the mobility of DSBs, alongside the repair defects. This reduced motion contributes to a reduction in the ability of a subset of DSBs to form clusters within the nucleus. The authors propose that this reduced clustering limits HR efficiency by decreasing the local concentration of factors that process DNA ends. Given that relatively few DSBs form clusters, another plausible scenario is that actin-dependent DSB movement and DNA-end processing are parallel events, and are not functionally linked.

The two new studies provide a framework for further investigations into the role of nuclear polymerized actin in the cell's response to DNA damage. Although the key biological players that support the nuclear functions of actin have been identified, the physical connections that attach damaged chromatin to nuclear actin structures and to motor proteins are unknown. The next step should be to use super-resolution microscopy and other imaging tools to fully characterize

© 2018 Springer Nature Limited. All rights reserved.

actin structures in the nucleus and to discover how the architecture of chromatin and the nucleus affects the formation and remodelling of actin polymerization. Further studies are also required to unravel the mechanism by which polymerized actin facilitates chromatin mobility and DNA-end processing. Given that genomic instability is a major contributor to the development of cancer, working out how cells use actin to safeguard their genome will have implications for our understanding of the basic principles of cancer aetiology. ■

Vassilis Roukos *is at the Institute of Molecular Biology, 55128 Mainz, Germany.*
e-mail: v.roukos@imb-mainz.de

1. de Lanerolle, P. & Serebryannyy, L. *Nature Cell Biol.* **13**, 1282–1288 (2011).
2. Belin, B. J. & Mullins, R. D. *Nucleus* **4**, 291–297 (2013).
3. Caridi, C. P. *et al. Nature* **559**, 54–60 (2018).
4. Schrank, B. R. *et al. Nature* **559**, 61–66 (2018).
5. Chiolo, I. *et al. Cell* **144**, 732–744 (2011).
6. Tsouroula, K. *et al. Mol. Cell* **63**, 293–305 (2016).
7. Jakob, B. *et al. Nucleic Acids Res.* **39**, 6489–6499 (2011).
8. Belin, B. J., Lee, T. & Mullins, R. D. *eLife* **4**, e07735 (2015).
9. Andrin, C. *et al. Nucleus* **3**, 384-395 (2012).
10. Chuang, C. H. *et al. Curr. Biol.* **16**, 825–831 (2006).
11. Dundr, M. *et al. J. Cell Biol.* **179**, 1095–1103 (2007).
12. Cho, N. W., Dilley, R. L., Lampson, M. A. & Greenberg, R. A. *Cell* **159**, 108–121 (2014).
13. Aymard, F. *et al. Nature Struct. Mol. Biol.* **4**, 366–374 (2014).
14. Massaad, M. J., Ramesh, N. & Geha, R. S. *Ann. NY Acad. Sci.* **1285**, 26–43 (2013).

This article was published online on 20 June 2018.

STRUCTURAL BIOLOGY

Picturing a key brain protein

Proteins called $GABA_A$ receptors have a pivotal role in neuronal inhibition, and are targets of several drugs. Structures of the most abundant $GABA_A$ receptor in humans are now reported, and will aid future drug discovery. SEE ARTICLE P.67

ERWIN SIGEL

The daytime biological activity of humans is based on a delicate equilibrium between excitatory and inhibitory signalling in the brain. The neurotransmitter molecule that mediates the main inhibitory signal is γ-aminobutyric acid (GABA), which binds to and activates ion channels, including a family known as $GABA_A$ receptors[1] found in the cell membranes of neurons. The activity of these receptors is modulated by many drugs[2], including sedatives, anxiolytics and sleeping pills. $GABA_A$ receptors can be assembled from several types of subunit, and the most abundant $GABA_A$ receptors in the adult human brain consist of two α1 subunits, two β2 subunits and one γ2 subunit. On page 67, Zhu *et al.*[3] present two structures of this receptor isoform in complex with GABA and a drug molecule known as flumazenil, providing much-needed insight that will aid future drug-discovery efforts.

The $GABA_A$ receptors that mediate fast signalling are found in several locations, most importantly at the synaptic junctions between neurons. Once activated by GABA, the channels in the receptors open, and conduct chloride ions through their pores. Benzodiazepines are a widely used class of drug that acts at some subtypes of $GABA_A$ receptor by binding to a specific site[4] distinct from that at which GABA binds. These compounds do not induce receptor activity on their own, but increase activity triggered by GABA, a process known as positive allosteric modulation.

Zhu *et al.* prepared the receptors for their study by expressing them in human embryonic kidney (HEK) cells in the presence of GABA and flumazenil, which is an antagonist for the benzodiazepine binding site (that is, it blocks the site). The authors used cryo-electron microscopy to obtain two high-resolution structures of the synaptic α1β2γ2 $GABA_A$ receptor (see Fig. 1 of the paper[3]), which they propose are in non-conducting, desensitized states — closed states of the channel that occur when the receptors have had prolonged exposure to GABA. Crucially, the structure confirms that the number of each subunit type and the arrangement of the subunits in the assembled pentameric receptor correspond to what had previously been postulated[1].

The structures reveal differences between the GABA binding sites (which are found at the two β2–α1 subunit interfaces in the receptor), the benzodiazepine binding site (which is found at the α1–γ2 interface, at an equivalent position to the GABA binding site), and equivalent sites at the α1–β2 and γ2–β2 interfaces to which neither GABA nor benzodiazepines bind (Fig. 1). These differences help to explain why the different sites bind (or do not bind) GABA or benzodiazepines. Both the γ2–β2 and α1–β2 sites are potential targets for drug discovery. A compound that binds at the α1–β2 interface has already been reported[5], but no compounds have yet been found for γ2–β2. The new receptor structures will provide a

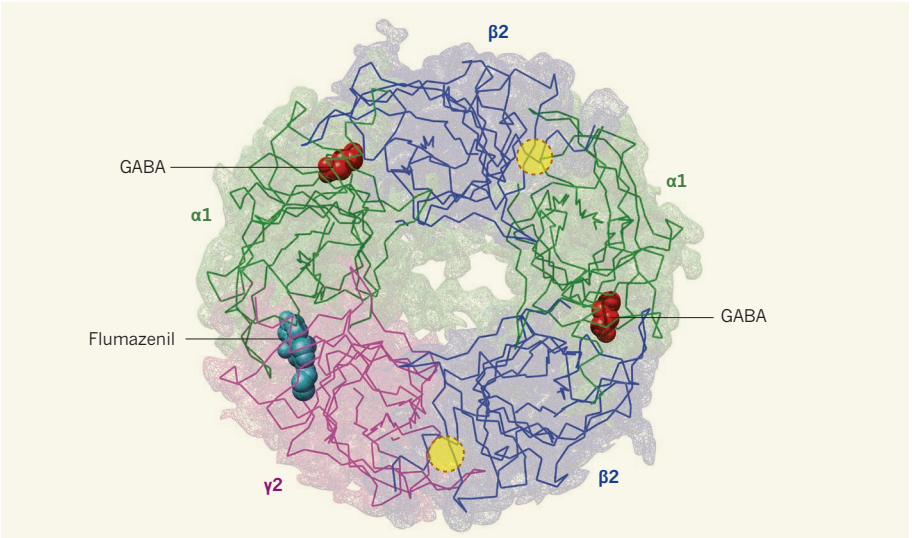


Figure 1 | Structure of the α1β2γ2 $GABA_A$ receptor viewed from the synapse. $GABA_A$ receptors are involved in neuronal inhibition and are targeted by several drugs. They form channels in the membranes of neurons, at different locations, most importantly at the synaptic junctions that connect pairs of neurons. Zhu *et al.*[3] report the structure of the most abundant $GABA_A$ receptor found in adult humans, which consists of two α1 subunits, two β2 subunits and one γ2 subunit. A molecule of γ-aminobutyric acid (GABA) — the naturally occurring ligand that activates the receptors — is bound at each of the GABA-binding sites, which are located at the two β2–α1 subunit interfaces. The drug flumazenil is bound at the benzodiazepine site at the α1–γ2 subunit interface (the site at which drug molecules from the benzodiazepine class bind). The γ2–β2 and α1–β2 interfaces contain pockets (yellow circles) that are at equivalent positions to the GABA and benzodiazepine sites, and which might be suitable targets for drug discovery. (Adapted from Fig. 2c of ref. 3.)

© 2018 Springer Nature Limited. All rights reserved.

template for molecular-modelling studies aimed at discovering more compounds that bind at the different interfaces, or to any of the many other pockets present in the receptor.

The extracellular region of the receptor forms a wide chamber. Zhu *et al.* find that most of the space in this chamber is occupied by complex carbohydrate chains attached to some of the subunits, leaving relatively little space for ion permeation. If the same is true for receptors expressed in neurons, rather than in HEK cells, then this observation is mechanistically highly relevant. Ions might also enter the channel through openings found at the subunit interfaces. A structure of the α1β1γ2 $GABA_A$ receptor was recently published[6] on a preprint server, and the authors of that paper suggest that the attachment of carbohydrates to the α1 subunit has a key role in receptor assembly.

Understanding the variations between benzodiazepine binding sites found in different $GABA_A$-receptor isoforms might help researchers to find compounds that act selectively at just one isoform. This could allow researchers to realize the hope of developing drugs that produce desirable therapeutic effects (such as reducing anxiety) without the unwanted side effects (such as sedation) caused by existing drugs that target several isoforms of $GABA_A$ receptors. How diazepam — the archetypal benzodiazepine drug — binds to benzodiazepine sites is of particular interest. Computational studies[7] suggest that there are three possible binding modes (BM I–III) of benzodiazepines, of which BM I was proposed to actually occur. By contrast, an alternative approach has been used to identify the amino-acid residues in α1β2γ2 $GABA_A$ receptors that come into direct contact with diazepam[8], and this pointed instead to BM II. The information from that study was used to identify high-affinity ligands for the benzodiazepine site.

In Zhu and colleagues' structures, the position of flumazenil in the benzodiazepine site is similar to BM I. The authors suggest that docking of diazepam to the observed site would be possible through a binding mode similar to BM I, but that diazepam docking similar to BM II would lead to clashes between the molecule and the receptor. It should be noted, however, that the receptor conformations stabilized by flumazenil must be different from that stabilized by diazepam, so the idea that diazepam would dock into the reported structure in a similar way to flumazenil is questionable. A recent study[9] has shown that benzodiazepines whose structures are similar to that of flumazenil use BM I, whereas those similar to diazepam use BM II. Thus, the positioning of diazepam postulated by Zhu *et al.* will probably need to be revised.

In structural biology, there is always the possibility that the protein structure determined does not entirely correspond to a conformation adopted in nature. This is illustrated by one of Zhu and colleagues' structures (described as conformation A), in which the transmembrane part of the γ2 subunit is squeezed into the receptor's pore. This conformation is unlikely to occur in nature, and might have been caused by the type of detergent that Zhu *et al.* used to replace the natural membrane environment in their microscopy experiments. The water-soluble domain of the protein is not subject to this problem.

Proteins are dynamic structures that assume distinct conformational states. In each state, large parts of the proteins vibrate much like wobbly puddings. The binding of molecules to allosteric sites in proteins (such as the benzodiazepine site in $GABA_A$ receptors) can stabilize different conformational states. This explains how molecules can act as positive or negative allosteric modulators, or as antagonists. $GABA_A$ receptors exist in at least four conformational states: closed; ligand-bound but not open (known as the pre-activated state); open; and desensitized. Therefore, insight into allosteric modulation cannot be gained from structural studies alone, because protein structures are caught in static conformations. Structural-biology methods will need to be combined with other approaches to probe the dynamic properties of $GABA_A$ receptors.

Nevertheless, Zhu and colleagues' structures could well be the key to addressing some of the crucial issues in the study of $GABA_A$ receptors. For example, the work might help to unravel the numerous conformations of the α1β2γ2 $GABA_A$ receptor, and the differences in the benzodiazepine binding site between different isoforms of $GABA_A$ receptors, as well as providing information on all the possible modulatory sites. It is to be hoped that this will, in turn, lead to the development of improved drugs that target these receptors. ■ SEE EDITORIAL P.5

Erwin Sigel *is at the Institute of Biochemistry and Molecular Medicine, Medical Faculty, University of Bern, 3010 Bern, Switzerland.*
e-mail: sigel@ibmm.unibe.ch

1. Sigel, E. & Steinmann, M. E. *J. Biol. Chem.* **287**, 40224–40231 (2012).
2. Sieghart, W. *Adv. Pharmacol.* **72**, 53–96 (2015).
3. Zhu, S. *et al. Nature* **559**, 67–72 (2018).
4. Sigel, E. & Ernst, M. *Trends Pharmacol. Sci.* **39**, 659–671 (2018).
5. Ramerstorfer, J. *et al. J. Neurosci.* **31**, 870–877 (2011).
6. Phulera, S. *et al.* Preprint at bioRxiv http://doi.org/crcz (2018).
7. Richter, L. *et al. Nature Chem. Biol.* **8**, 455–464 (2012).
8. Middendorp, S. J., Puthenkalam, R., Baur, R., Ernst, M. & Sigel, E. *ACS Chem. Biol.* **9**, 1854–1859 (2014).
9. Elgarf, A. A. *et al. ACS Chem. Biol.* https://doi.org/10.1021/acschembio.8b00144 (2018).

NEUROSCIENCE

Social isolation's molecular signature

Extended social isolation causes debilitating effects in social mammals such as humans. A study in mice shows that the gene *Tac2* is upregulated throughout the brains of socially isolated animals, driving massive behavioural changes.

NOGA ZILKHA & TALI KIMCHI

Even the toughest prisoners fear solitary confinement. There is a growing awareness across the globe that we are facing an epidemic of loneliness. Prolonged social isolation and loneliness can lead to many profound physiological and neuropsychiatric conditions, including depression and heart disease, and to increased mortality rates[1]. In the United States, more than 50% of people over the age of 60 experience loneliness[2], and the United Kingdom has appointed a government minister to tackle the issue of loneliness. But the biological mechanisms underlying the effects of social isolation are poorly understood. Writing in *Cell*, Zelikowsky *et al.*[3] reveal a signalling mechanism that acts in several brain regions in mice to drive some of the harmful effects of the stress caused by chronic social isolation.

The authors examined the effects of two weeks of social isolation on the brains and behaviour of male mice (equivalent to more than a year in these conditions for humans[4]). First, the researchers used an array of behavioural tests to compare mice kept in isolation with control mice that had been housed in groups. These assays revealed widespread effects. Compared to control animals, isolated mice showed enhanced aggression and hypersensitivity to diverse stressful stimuli. For example, the socially isolated mice responded more aggressively to an unfamiliar mouse placed in their cage. In another assay, the researchers presented mice with a dark circle that loomed overhead, simulating an approaching predator. Control animals froze in response to the threat, but moved normally after the stressful stimulus was removed, whereas isolated mice remained frozen long after the apparent threat was removed.

Next, Zelikowsky *et al.* investigated the brain mechanisms underlying this behaviour. In a

© 2018 Springer Nature Limited. All rights reserved.

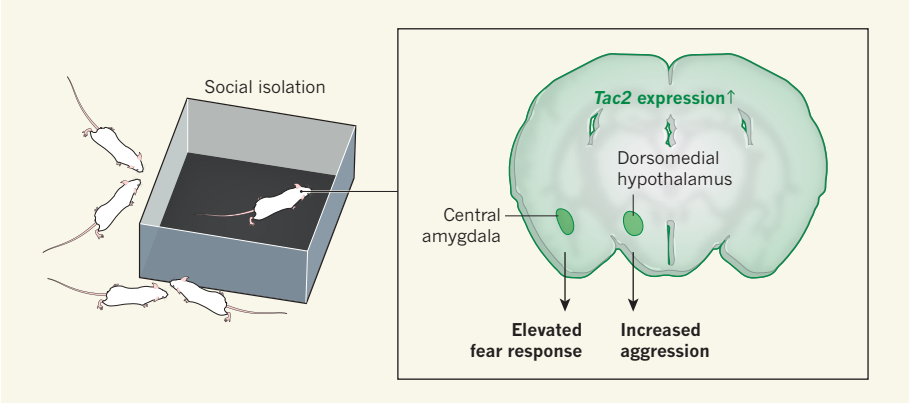


Figure 1 | The gene *Tac2* mediates various effects of social isolation in mice. Zelikowsky *et al.*[3] investigated how two weeks of isolation affected the brains and behaviour of male mice. They found that *Tac2* expression is upregulated throughout the brain, and that the gene's upregulation in particular areas — including the central amygdala and dorsomedial hypothalamus — led to specific changes in the animals' social behaviour and in their response to various stressful stimuli.

previous study of fruit flies, the same group had identified the gene *Tac* as essential for the regulation of aggression induced by social isolation[5]. Rodents have two versions of *Tac*, which are expressed in various brain regions, including regions associated with social behaviour, anxiety and emotions. Using several independent methods, Zelikowsky and colleagues now found a massive increase in the expression of *Tac2* throughout the brain following social isolation.

The gene *Tac2* encodes a protein called neurokinin B (NkB), which binds specifically to the receptor Nk3R. The researchers performed a series of experiments to alter NkB signalling in the brain. First, they systemically inhibited NkB signalling in isolated male mice using a drug called osanetant, which inhibits the activity of Nk3R. Administration of osanetant, either throughout the social-isolation period or 20 minutes before behavioural testing, substantially reduced the effects of social isolation on behaviour. Next, the authors genetically upregulated *Tac2* expression and simultaneously activated *Tac2*-expressing neurons in group-housed animals, using specially designed viruses that were injected intravenously but could cross the blood–brain barrier to reach the brain. They found that this genetic manipulation led to group-housed mice behaving in a similar way to those that had been isolated.

Finally, Zelikowsky *et al.* locally manipulated *Tac2* expression and NkB signalling, by injecting either osanetant or viruses to downregulate *Tac2* expression or inhibit the activity of *Tac2*-expressing neurons, into particular locations in the brain. These experiments enabled the authors to attribute specific behaviours to regulation of *Tac2* in specific brain regions. The main social effect of isolation — enhanced aggression towards an intruder — was controlled by *Tac2* in the dorsomedial hypothalamus. By contrast, acute and persistent stress responses were regulated primarily by *Tac2* in the central amygdala (Fig. 1).

This work opens a gateway to much future research. First and foremost, it will be interesting to determine whether *TAC3*, the human equivalent of *Tac2*, is involved in mediating the effects of loneliness and social isolation in people. To our knowledge, *TAC3* has not yet been directly associated with sociality or social behaviour of any kind in humans. However, it is expressed in the human brain and has shown abnormal gene-expression levels in children with autism-spectrum disorder[6], which profoundly affects social interaction. The systemic manipulations presented in Zelikowsky and colleagues' paper could be rapidly applied to humans, because osanetant and other NkB inhibitors have already been tested in clinical trials. These drugs could potentially treat antisocial disorders induced by isolation, as well as mood and anxiety disorders.

Although most of their experiments focused on male mice, Zelikowsky *et al.* found upregulation of *Tac2* in response to social isolation in both males and females. Sex differences in response to stress and isolation are well documented, and are usually conserved across species[7]. It will therefore be interesting to test whether the roles of *Tac2* in mediating the effects of social isolation in females are similar to or different from those in males.

The need for social interactions and the response to social isolation can differ enormously between and within species. Mice and humans, for example, are typically considered to be highly social creatures[8]. When their social needs are not filled, they can experience debilitating outcomes[1,9]. Some species (and individuals within a species), however, are more solitary, or even avoid social interactions[10]. Such species or individuals might harbour neuronal mechanisms that are adapted to the lack of social interaction. Whether or not members of the *Tac* gene family act differently in solitary individuals or species compared to how they do in more-social individuals or species remains to be determined.

Finally, one has to wonder: to what extent



50 Years Ago

Thirty years ago, those of us who argued that hunger and malnutrition existed and would get worse unless research on agriculture and contraception got very much more support, were variously labelled atheists, communists, dreamers or idealists according to the prejudices of the critic. But facts are stubborn, and awareness of the need for more food is now widespread. Consequently, we are deluged with books and symposia on what could and should be done to balance food and population … Optimists tend to overlook the fact that, unless a population control policy depends on a dictatorial act along Herodian lines, it will not be effective until people have learnt the rudiments of biology and hygiene.
From *Nature* 6 July 1968

100 Years Ago

Anthrax is an acute, infective disease of man and animals and is caused by the anthrax bacillus … In order to prevent the disease in dangerous trades working with possibly infected animal material it would, at first sight, appear to be a simple thing to disinfect the infected material. In practice, however, this is found to be exceedingly difficult on account of the truly enormous powers of resistance of the spore … complete success can now be attained without risk to the workers and without damage to the material disinfected. The main feature of the process is a preliminary treatment in which material is submitted to the action of a warm solution of soap and water containing alkali … This causes softening and disintegration of any infected blood-clots, and the spores are laid bare for the subsequent destroying process … ten million pounds of infected wool can be effectively disinfected for something less than 0 824 penny per pound weight.
From *Nature* 4 July 1918

© 2018 Springer Nature Limited. All rights reserved.

can we rely on a mouse model of social isolation to truly examine the underlying mechanisms of human loneliness? After all, loneliness and mental isolation are subjective, and a person might feel alone even when surrounded by other people. The traits exhibited by mice under prolonged social isolation greatly resemble those found in humans experiencing solitary confinement, so these animals do provide a good model for studying this process. What we currently lack are relevant animal models for other forms of human loneliness, such as social withdrawal or antisocial personality disorder. Expanding our research toolbox — for example, by studying various species, including non-social and community-living animals, as well as humans — might bring us closer to understanding the biology of human loneliness. ■

Noga Zilkha *and* **Tali Kimchi** *are in the Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel. e-mail: tali.kimchi@weizmann.ac.il*

1. Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T. & Stephenson, D. *Perspect. Psychol. Sci.* **10**, 227–237 (2015).
2. Gerst-Emerson, K. & Jayawardhana, J. *Am. J. Public Health* **105**, 1013–1019 (2015).
3. Zelikowsky, M. *et al. Cell* **173**, 1265–1279 (2018).
4. Dutta, S. & Sengupta, P. *Life Sci.* **152**, 244–248 (2016).
5. Asahina, K. *et al. Cell* **156**, 221–235 (2014).
6. Martin, J. *et al. J. Am. Acad. Child Adolescent Psychiatry* **53**, 761–770 (2014).
7. Bale, T. L. & Epperson, C. N. *Nature Neurosci.* **18**, 1413 (2015).
8. Insel, T. R. & Fernald, R. D. *Annu. Rev. Neurosci.* **27**, 697–722 (2004).
9. House, J., Landis, K. & Umberson, D. *Science* **241**, 540–545 (1988).
10. Rehan, S. M. & Toth, A. L. *Trends Ecol. Evol.* **30**, 426–433 (2015).

This article was published online on 20 June 2018.

ASTRONOMY

General relativity passes another test

Einstein's theory of gravity — the general theory of relativity — is based on the principle that all objects accelerate identically in an external gravitational field. A triple-star system provides a stringent test of this principle. SEE LETTER P.73

CLIFFORD M. WILL

All bodies in a given gravitational field are thought to fall with the same acceleration. This idea, known as the equivalence principle, is central to our understanding of gravitational physics. It was promoted by thinkers ranging from the sixth-century scholar John Philoponus to Galileo; it is the founding principle of Albert Einstein's general theory of relativity, and was famously demonstrated when Apollo astronaut David Scott dropped a hammer and a feather on the Moon and saw that they hit the lunar surface at the same time. For decades, experimentalists have verified the equivalence principle using exquisitely delicate instruments. Now, on page 73, Archibald *et al.*[1] report the results of a remarkable test of the principle, in which the falling objects are two stellar remnants: a neutron star and a white dwarf.

A spinning neutron star that emits a beam of electromagnetic radiation is known as a pulsar. The emission seems to pulse because it can be seen only when the beam is pointing towards Earth. The pulses are so regular that variations in their observed period can be readily interpreted as being due to the gravitational tug of another astronomical body on the pulsar. Such variations have been used to discover more than 220 binary systems containing a neutron star, and a handful of pulsars that have associated planets (go.nature.com/2mslfj4).

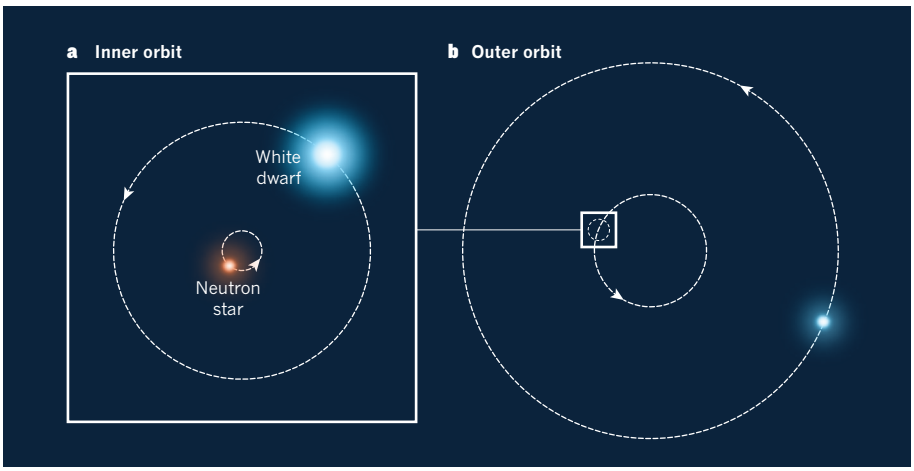


Figure 1 | Triple-star system. In 2014, astronomers reported a system that contains three stellar remnants: a neutron star and two white dwarfs[2]. **a**, The neutron star is in a close 1.6-day orbit with one of the white dwarfs. **b**, This pair of objects is itself in a 327-day orbit with the other white dwarf. Archibald *et al.*[1] report no evidence of a deformation of the inner orbit, which would be expected if there were a difference between the accelerations of the neutron star and the inner white dwarf towards the outer white dwarf. The results provide support for Einstein's theory of gravity — the general theory of relativity.

In 2014, astronomers reported a pulsar that is unusual because it has two stellar companions[2] (Fig. 1). The neutron star, weighing 1.4 solar masses, is in a close 1.6-day orbit with a 0.2-solar-mass white dwarf. This pair of objects is itself in a 327-day orbit with a 0.4-solar-mass white dwarf. The inner and outer orbits are nearly circular and exist in almost exactly the same plane.

If the neutron star and the inner white dwarf were to fall with different accelerations towards the outer white dwarf, there would be a tiny deformation of the inner orbit. Archibald and colleagues report an analysis of approximately six years of data showing no evidence of such a deformation. The accelerations of the two bodies differ by no more than 2.6 parts per million, in agreement with the equivalence principle.

Tests of this principle have a long heritage. In the late nineteenth century, the Hungarian physicist Roland von Eötvös devoted years to verifying that the accelerations of various laboratory materials in Earth's gravitational field differ by less than a few parts per billion[3]. His modern-day successors, the Eöt-Wash group[4] in Seattle, Washington, pushed this bound to parts per 10^{13}. And, in 2017, data from the French space mission MICROSCOPE[5] moved the goalpost by a further factor of ten.

Given that a typical object in a physics lab consists of a swarm of elementary particles and their associated fields and energies, it is quite extraordinary that the responses of different materials to gravity should be so similar. In Einstein's unique imagination, there was a reason: gravity is not a force that acts on all of these particles in some fantastically fine-tuned manner, but is simply an effect of space-time geometry. The constituents of matter follow universal paths in a space-time that is curved by massive bodies, such as Earth or the Sun.

But does gravitational energy act in the same way as matter? The small objects used in lab experiments do not contain enough

© 2018 Springer Nature Limited. All rights reserved.

gravitational energy to answer this question, but planets and stars do. With self-gravity in the picture, a concept called the strong equivalence principle comes into play. This principle singles out the general theory of relativity from its competitors. In Einstein's theory, all bodies — hammers, feathers, planets, neutron stars, white dwarfs and even black holes — fall with the same acceleration. But in most alternative theories of gravity, such as scalar–tensor theories[6], the equivalence principle is violated for bodies that have self-gravity.

For almost 50 years, researchers have measured how long it takes for laser pulses to make the round trip from Earth to the Moon and back — a technique known as lunar laser ranging. Analyses of these data[7,8] have verified the strong equivalence principle, by showing that the accelerations of the two bodies towards the Sun differ by no more than a few parts per 10^{13}. Because about 5 parts in 10^{10} of Earth's mass is gravitational energy[9], this result implies that the accelerations of gravitational energy and matter differ by less than a few parts per 10^{4}.

Archibald and colleagues' study breaks new ground because the gravitational energy inside a neutron star can account for as much as 20% of the body's mass[10]. The authors' results therefore imply that the accelerations of gravitational energy and matter differ by no more than a few parts per 10^{5} — a tenfold improvement over the bound from lunar laser ranging.

More importantly, the authors have provided what is known as a strong-field test of general relativity. Unlike the Solar System, for which Einstein's theory predicts only small deviations from Newton's theory of gravity, the motion of a neutron star in a gravitational field invokes full general relativity in all its complex glory. Einstein's theory passes this strong-field test with flying colours.

Because general relativity predicts a null effect, the grading is a simple pass or fail. But for alternative theories, invoking strong-gravity effects substantially complicates the interpretation of the results. Archibald *et al.* demonstrate this complexity using scalar–tensor theories as an example. For these theories, the interpretation of the results depends on the internal structure assumed for the neutron star and on the values chosen for quantities known as coupling constants. The authors show that their results improve on certain pre-existing constraints on the parameters that govern these theories — some arising from Solar System measurements and some from data on binary systems containing a pulsar. Although the theories are not completely quashed, their hopes for validity have been made that much fainter. ■

Clifford M. Will *is in the Department of Physics, University of Florida, Gainesville, Florida 32611, USA.*
e-mail: cmw@phys.ufl.edu

1. Archibald, A. M. *et al. Nature* **559**, 73–76 (2018).
2. Ransom, S. M. *et al. Nature* **505**, 520–524 (2014).
3. Eötvös, R. V., Pekár, D. & Fekete, E. *Ann. Phys. (Leipz.)* **373**, 11–66 (1922).
4. Wagner, T. A., Schlamminger, S., Gundlach, J. H. & Adelberger, E. G. *Class. Quantum Grav.* **29**, 184002 (2012).
5. Touboul, P. *et al. Phys. Rev. Lett.* **119**, 231101 (2017).
6. Brans, C. & Dicke, R. H. *Phys. Rev.* **124**, 925–935 (1961).
7. Williams, J. G., Turyshev, S. G. & Boggs, D. H. *Int. J. Mod. Phys. D* **18**, 1129–1175 (2009).
8. Hofmann, F., Müller, J. & Biskupek, L. *Astron. Astrophys.* **522**, L5 (2010).
9. Will, C. M. *Theory and Experiment in Gravitational Physics, Revised Edition* 174 (Cambridge Univ. Press, 1993).
10. Shapiro, S. L. & Teukolsky, S. A. *Black Holes, White Dwarfs and Neutron Stars: The Physics of Compact Objects* 241 (Wiley, 1986).

DEVELOPMENTAL BIOLOGY

Fat gets a brake

Single-cell transcriptional profiling of stem and progenitor cells in fat tissue identifies distinct cell subpopulations, one of which inhibits fat growth by signalling to neighbouring cells. SEE LETTER P.103

DAVID A. GUERTIN

Fat tissue has a remarkable capacity for growth. It can expand in two ways: by increasing the size of individual fat cells (adipocytes), and by making adipocytes from progenitor cells through the process of adipogenesis. Fat is essential for metabolic fitness, but having too much fat in the wrong places can be harmful. Obesity is a precursor to serious medical conditions such as type 2 diabetes, cardiovascular disease and cancer, which are ravaging health-care systems worldwide. Key to combating obesity is understanding how mature adipocytes develop from precursor cells, but the identity of these precursors has so far been elusive. On page 103, Schwalie *et al.*[1] define three populations of fat-precursor cell, one of which unexpectedly functions to suppress adipocyte production.

The origin of the body's adipocytes has been a mystery, complicated by the fact that fat tissues, called depots, contain many cell types other than adipocytes. In addition, adipocytes at different anatomical locations originate from different early embryonic precursors[2]. There is also metabolic variation between fat depots around the body, and even between adipocytes in the same depot[3–5].

A pool of adipocyte stem and progenitor cells (ASPCs) can be isolated from fat depots using a technique called fluorescence-activated cell sorting (FACS), which separates cells on the basis of specific cell-surface proteins[6,7]. Many of the surface markers currently used to isolate ASPCs through FACS were selected because they distinguish stem-like cells from other tissues. However, the pools isolated when trying to obtain ASPCs using this method contain a mixture of cell types, and the molecular profiles of the true stem and progenitor cells within the mix has remained largely undefined.

Schwalie and colleagues started with a mixture of 208 ASPCs isolated from mouse fat tissue using FACS. They performed single-cell RNA sequencing to determine which genes are expressed in each cell. Using computer algorithms to group the cells according to their gene-expression profiles, the authors discovered that at least three distinct subpopulations exist within the ASPC pool. They then confirmed the existence of these three subpopulations using an alternative cell-isolation strategy combined with a different method of grouping ASPCs by their gene-expression signatures.

Most of the ASPCs fell into two of three groups discovered by the authors. The first group, designated P1, expressed high amounts of stem-cell markers. The second group, designated P2, expressed many genes that regulate the early steps of adipocyte formation. But it was the smallest group, P3, representing less than 10% of the cell population, that drew the authors' attention. Unlike the other groups, P3 cells did not form mature adipocytes when

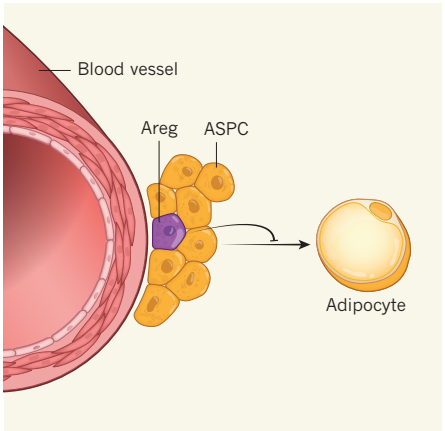


Figure 1 | A cell population that inhibits fat growth. A pool of adipocyte stem and progenitor cells (ASPCs) that is found around blood vessels gives rise to fat cells, which are called adipocytes. Schwalie *et al.*[1] identified a subpopulation of cells within the ASPC pool, dubbed Aregs. These cells release signals to inhibit the formation of adipocytes from ASPCs.

© 2018 Springer Nature Limited. All rights reserved.

induced to differentiate in a cell-culture dish. In addition, removing P3 cells from the ASPC pool improved the ability of the other cells in the dish to differentiate into adipocytes.

These data suggest that the low-abundance P3 cells inhibit adipogenesis. Schwalie and colleagues named these cells adipogenesis regulators (Aregs), and confirmed the cells' function *in vivo*. First, they transplanted two mixtures of ASPCs — one lacking Aregs, the other containing the entire cell pool — into mice. Each mouse received both mixtures, one on each side of the body. Next, the researchers fed the mice a high-fat diet to induce adipogenesis. Over a few weeks, the implanted cell mixture lacking Aregs grew many more adipocytes than the other mixture did, indicating that Aregs inhibit fat growth. The authors also showed that Aregs exist in human fat, implying that fat-development mechanisms are conserved between mice and humans.

How do Aregs inhibit adipogenesis? Schwalie *et al.* found that the cells reside near the blood vessels of fat tissues in mice, a location that was previously proposed as the site of adipocyte precursor cells[8] (Fig. 1). Next, the authors investigated whether Aregs signal to neighbouring cells through physical contact or by sending chemical (paracrine) signals to nearby cells. Co-culture experiments, in which the authors placed a barrier permeable to small molecules between the Aregs and their target cells, revealed that direct contact is not needed for Aregs to influence fat-cell formation, indicating that the signal is paracrine.

To identify candidate signalling molecules, the researchers inactivated genes that are highly expressed in Aregs. They found that the gene *Rtp3* needed to be turned on to enable Aregs to send their inhibitory signals. Little is known about the Rtp3 protein, and it is not obvious how it works in this context. This is an area ripe for future study, because modulating the signals released by Aregs could have therapeutic potential for controlling fat growth.

Schwalie and colleagues' findings are exciting for several reasons. First, although high variation between ASPC subpopulations had been predicted, this study fills a major gap by adding molecular details to our understanding of that variability. Second, the authors use state-of-the-art technology for single-cell gene-expression profiling, enabling them to identify a regulatory cell type that would have been difficult to predict on the basis of previous studies. It is to be hoped that this study will stimulate other work aimed at elucidating the organization of adipogenesis (the hierarchy of cells that regulate the formation of fat), as has been achieved for blood-cell lineages[9].

The current study adds to the mounting evidence that paracrine signals help to remodel stem- and progenitor-cell function[10], and opens up several avenues for future research. For instance, what is the anti-adipogenic signal, and how does Rtp3 help to stimulate Aregs to produce it? Genetic or age-related differences in Areg number or function might contribute to body-fat patterning or the propensity to become obese, and these possibilities should also be explored.

It will be interesting to determine whether the ASPC pool can be divided into further subpopulations with more-specific functions. Tracing the *in vivo* fates of these different subpopulations would be a powerful strategy for picking apart which cells become adipocytes and which become fat-supporting cells. Finally, perhaps one of the most interesting questions raised by the study is whether there is a true adult adipocyte stem cell, which, by definition, would be capable of producing both committed adipocyte progenitors and more adipocyte stem cells.

As the devastating human costs of obesity-related conditions rise, research and health-care professionals must meet the challenge with breakthroughs in medical management and care. This will require a better understanding of adipogenesis, and Schwalie and colleagues' work has pointed to a new way of advancing knowledge in this important area. ■

David A. Guertin *is at the University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.*
e-mail: david.guertin@umassmed.edu

1. Schwalie, P. C. *et al. Nature* **559**, 103–108 (2018).
2. Sanchez-Gurmaches, J., Hung, C.-M. & Guertin, D. A. *Trends Cell Biol.* **26**, 313–326 (2016).
3. Lynes, M. D. & Tseng, Y.-H. *Ann. NY Acad. Sci.* **1411**, 5–20 (2018).
4. Schoettl, T., Fischer, I. P. & Ussar, S. *J. Exp. Biol.* **221**, jeb162958 (2018).
5. Rosen, E. D. & Spiegelman, B. M. *Cell* **156**, 20–44 (2014).
6. Hepler, C., Vishvanath, L. & Gupta, R. K. *Genes Dev.* **31**, 127–140 (2017).
7. Berry, R., Jeffery, E. & Rodeheffer, M. S. *Cell Metab.* **19**, 8–20 (2014).
8. Tang, W. *et al. Science* **322**, 583–586 (2008).
9. Orkin, S. H. & Zon, L. I. *Cell* **132**, 631–644 (2008).
10. Kusuma, G. D., Carthew, J., Lim, R. & Frith, J. E. *Stem Cells Dev.* **26**, 617–631 (2017).

This article was published online on 20 June 2018.

TISSUE ENGINEERING

Living replacement heart valves remodelled

Bioengineered heart valves are a promising treatment for heart-valve disease, but often undergo mechanical failure when implanted. Computational modelling of the initial valve design has now improved their performance in sheep.

CRAIG A. SIMMONS

Heart-valve disease has been described as an emerging epidemic[1], owing to its worldwide prevalence, its potential to kill, and the lack of therapies for its prevention or treatment. Damaged and defective valves can be replaced with prosthetic ones, but the inability of prostheses to grow or adapt to change makes them a poor solution for young patients[2]. An alternative is a living replacement made from bioengineered tissue. But, so far, tissue-engineered heart valves (TEHVs) have failed because detrimental valve-tissue remodelling occurs *in vivo*, impairing normal function. Writing in *Science Translational Medicine*, Emmert *et al.*[3] address this problem using computational modelling to design a TEHV that remodels favourably after implantation.

A promising strategy for heart-valve tissue engineering is to grow tissues in the shape of a valve in the laboratory using cells and a degradable biomaterial, then remove the cells to leave an empty extracellular-matrix scaffold. After implantation in the heart, the tissue scaffold is populated by the recipient's cells, presumably from the blood and the adjacent blood-vessel wall. Some of these cells then transform into contractile cells that degrade the scaffold and replace it with new tissue, while pulling on the tissues to hold everything together, and speed up remodelling. In all animal studies so far, however, this process has led to excessive tissue production, which thickens and stiffens the valve's three leaflets to obstruct blood flow, or to excessive tissue contraction, which causes the leaflets to shorten and retract, preventing proper valve closure and allowing backflow of blood (Fig. 1a).

After valve repair in healthy hearts, the contractile cells typically enter a deactivated state called quiescence, or are cleared from the region by programmed cell death. But certain stimuli, such as biomechanical stresses[4], can cause the cells to persist and remain activated. This can lead to excessive collagen production and tissue contraction — a process known as fibrosis[4]. In the native valve, fibrosis leads to leaflet thickening or retraction, which is similar to the problem seen in TEHVs. Biomechanical activation of fibrosis can be a particular problem in heart valves that are subjected to abnormalities in stretch, compression and pressure changes as they open

© 2018 Springer Nature Limited. All rights reserved.

and close with each heartbeat[5].

Using computational modelling, the group behind the current study have previously shown[6] that the leaflets of conventional TEHVs are compressed in the radial direction (towards the centre of the valve) by blood pressure when the valve is closed. This type of compression, which does not occur in native valves, acts to shorten the leaflets. It is also associated with contractile-cell activation and fibrotic remodelling.

One way to address this problem would be to directly inhibit fibrosis. But Emmert *et al.* took an alternative approach, attempting to limit leaflet shortening by minimizing radial compressions, and to guide the remodelling process to prevent persistent contractile-cell activation. The authors used computational modelling to design TEHVs that initially had a non-physiological geometry predicted to minimize radial compression, and a large area of contact between the leaflets when the valve was closed. The computational models predicted that, after implantation, these valves would remodel into a stable geometry that mimicked the shape of native valves (Fig. 1b).

The authors tested this hypothesis by replacing heart valves with computationally inspired TEHVs in ten sheep. They assessed the valves' performance over one year. As the computational models predicted, the valves adopted a stable tissue architecture *in vivo*. The valves performed comparably to native valves in nine out of ten animals. The recipient cells that populated the TEHVs produced new collagen without leaflet thickening. Collagen fibres aligned around the circumference of the valve, although not to the same extent as in native valves. The leaflets shortened as the valves remodelled, but shortening stabilized six months after implantation and the leaflets remained in contact with one another, so valve function was not affected. Finally, although many recipient cells infiltrated the valves, few of these were activated contractile cells after one year. Together, these results demonstrate that *in vivo* TEHV remodelling can be guided towards stable physiological structure and function.

Perhaps the most striking aspect of this study is that the authors' predictive-modelling approach was successful even though they could not control the cell types that infiltrated the TEHVs. But although this approach worked well in healthy sheep, the cells that repopulate the tissue scaffold in people who have valve diseases or defects might respond differently to mechanical stimuli. Indeed, the authors' computational model predicted that the TEHVs would work best when cell contractility was low, which may not always be the case in humans. For example, contractile-cell activation is driven by inflammation and immune responses[7] that were largely absent in this study, but are often elevated in people with valve disease. In this context, excessive contractile-cell activation could lead to detrimental fibrosis and TEHV failure.

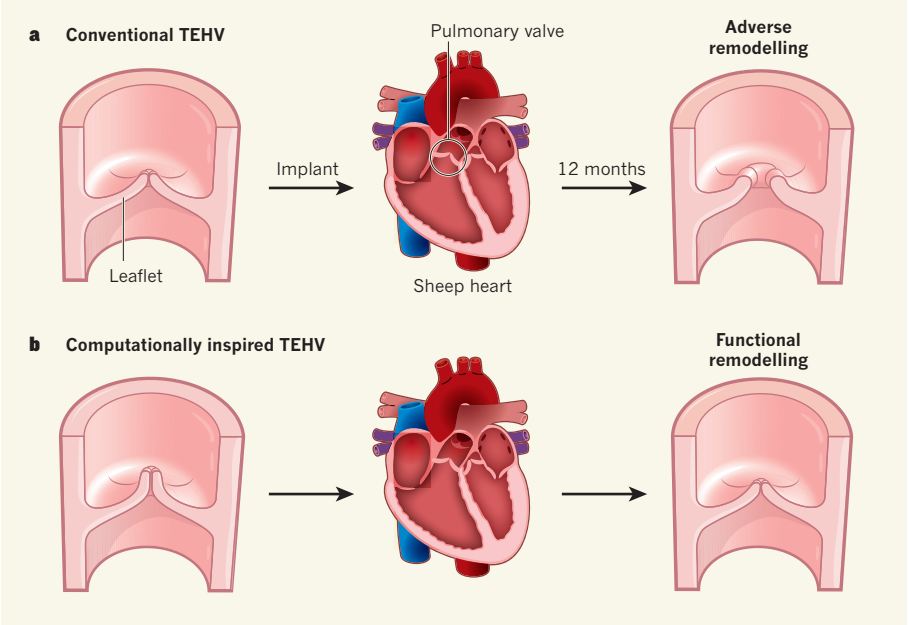


Figure 1 | Improving tissue-engineered heart valves using computational modelling. Heart valves can be replaced with alternatives made of living tissue. **a**, In conventional tissue-engineered heart valves (TEHVs), leaflets make small contacts with one another. But when these TEHVs are implanted into sheep hearts to replace the native pulmonary valve, they undergo adverse remodelling changes after 12 months. The leaflets become thicker and retract, leaving a hole through which blood leaks. **b**, Emmert *et al.*[3] have used computational modelling to improve the initial valve design. Their modelling predicted that a valve shape that minimized shortening of the leaflets under pressure (not depicted), and had large initial contact areas between leaflets, would remodel more favourably after implantation. Indeed, 12 months later, valve geometry and function were stable and comparable to that of native valves.

Moving forward, it will therefore be important to determine the robustness of a strategy that involves computational modelling based solely on mechanical influences, given the diversity of pro-fibrotic stimuli that can derail remodelling in patients. Successful TEHV solutions will probably need to combine computationally guided remodelling with complementary strategies to control detrimental fibrosis. For example, native valve cells express anti-fibrotic factors[8] that could potentially be delivered alongside a TEHV to help suppress adverse remodelling, particularly in the early post-implantation phase, when inflammation might be heightened.

It is also notable that the remodelled valves showed little development of the trilayered microstructure of native valves, in which the top, middle and underlying sections of each leaflet have different compositions and mechanical properties. This structure is thought to be essential to native-valve mechanics[5]. Longer-term follow-up will be necessary to determine whether Emmert and colleagues' TEHVs undergo further remodelling after 12 months to produce this microstructure, as has been observed in other TEHVs[9], or if normal function can be maintained without it. Longer-term follow-up will also reveal whether the recipient cells in the TEHVs — which had quiesced and stopped making tissue after one year — can be reactivated to make new tissue and allow the valve to grow, as is required for children.

Ultimately, TEHV cells must be activated to grow and repair as necessary, but then quiesce to prevent fibrosis. Emmert and colleagues' study clearly demonstrates the potential of a computational strategy to design TEHVs that can achieve this delicate balance on the basis of predicted mechanical and biological outcomes. The work also argues for further development of this approach to account for other factors in remodelling that could be predictably guided. ■

Craig A. Simmons *is in the Translational Biology & Engineering Program at the Ted Rogers Centre for Heart Research, and in the Department of Mechanical & Industrial Engineering and Institute of Biomaterials & Biomedical Engineering, University of Toronto, Toronto M5G 1M1, Canada.*
e-mail: c.simmons@utoronto.ca

1. d'Arcy, J. L. *et al. Eur. Heart J.* **37**, 3515–3522 (2016).
2. Parvin Nejad, S., Blaser, M. C., Santerre, J. P., Caldarone, C. A. & Simmons, C. A. *Adv. Drug Deliv. Rev.* **96**, 161–175 (2016).
3. Emmert, M. Y. *et al. Sci. Transl. Med.* **10**, eaan4587 (2018).
4. Tomasek, J. J., Gabbiani, G., Hinz, B., Chaponnier, C. & Brown, R. A. *Nature Rev. Mol. Cell Biol.* **3**, 349–363 (2002).
5. Sacks, M. S., Merryman, W. D. & Schmidt, D. E. *J. Biomech.* **42**, 1804–1824 (2009).
6. Loerakker, S., Argento, G., Oomens, C. W. J. & Baaijens, F. P. T. *J. Biomech.* **46**, 1792–1800 (2013).
7. Van Linthout, S., Miteva, K. & Tschöpe, C. *Cardiovasc. Res.* **102**, 258–269 (2014).
8. Blaser, M. C. *et al. Circ. Res.* **122**, 405–416 (2018).
9. Hoerstrup, S. P. *et al. Circulation* **102** (Suppl. III), III44–III49 (2000).

© 2018 Springer Nature Limited. All rights reserved.

REVIEW

https://doi.org/10.1038/s41586-018-0259-z

Structural insights into G-protein-coupled receptor allostery

David M. Thal[1]*, Alisa Glukhova[1], Patrick M. Sexton[1] & Arthur Christopoulos[1]*

G-protein-coupled receptors (GPCRs) are key cell-surface proteins that transduce external environmental cues into biochemical signals across the membrane. GPCRs are intrinsically allosteric proteins; they interact via spatially distinct yet conformationally linked domains with both endogenous and exogenous proteins, nutrients, metabolites, hormones, small molecules and biological agents. Here we explore recent high-resolution structural studies, which are beginning to unravel the atomic details of allosteric transitions that govern GPCR biology, as well as highlighting how the wide diversity of druggable allosteric sites across these receptors present opportunities for developing new classes of therapeutics.

G-protein-coupled receptors (GPCRs) are the largest superfamily of cell-surface receptors encoded by the human genome. They are pre-eminent environmental sensors and mediate signalling of a diverse range of extracellular activating ligands (agonists) including photons, ions, sensory stimuli, lipids, hormones, neurotransmitters, metabolites, peptides and proteins. GPCRs have evolved to transmit external environmental signals from one part of the protein to another. Therefore, they are intrinsically allosteric because GPCR signal transduction involves communication between spatially distinct yet conformationally linked binding sites[1]. In response to agonist binding to the cognate 'orthosteric' site, GPCRs undergo a conformational change to interact with intracellular transducers such as heterotrimeric G proteins and arrestins, ultimately resulting in integrated cellular responses. A small fraction of GPCRs already account for ~30% of US Food and Drug Administration (FDA)-approved medicines, highlighting their vital role in health and disease, high degree of tractability to drug discovery, and scope for further exploitation for leveraging understanding of GPCR biology[1,2]. This situation has developed rapidly within the last decade owing to breakthroughs enabled by the application of structural and biochemical studies of GPCRs, particularly advances in protein purification and engineering, lipid-based crystallography, X-ray diffraction, cryo-electron microscopy (cryo-EM) and computational biology[3]. All GPCRs share a common seven-transmembrane domain (7TMD) architecture and more variable extracellular domains (ECDs). There are now more than 200 reported GPCR structures (PDB entries) from more than 50 unique receptors, including full-length structures from members of classes A, B and F, and incomplete (7TMD or ECD) structures from class B, C and F GPCRs[3] (Supplementary Table 1). Collectively, these structural studies have provided new insights into the allosteric nature of GPCRs, including activation, modulation by exogenous and endogenous molecules, and dimerization (Box 1). This Review highlights some of the key insights and challenges for the field arising from this new information for understanding allosteric mechanisms in GPCR biology.

Allosteric transitions govern GPCR activation

The simplest GPCR-activation mechanism is one of conformational selection comprising two states, with an 'allosteric transition' mediating the change between inactive and active states. Agonists preferentially stabilize the active state, inverse agonists stabilize the inactive state, whereas 'neutral' antagonists show similar affinities for both states but can block the actions of both agonists and inverse agonists if they interact via an overlapping site. Rhodopsin is a prototypical receptor that exemplifies this mechanism, and numerous structures have been reported that capture key states of rhodopsin (Supplementary Table 1). Rhodopsin is unusual among GPCRs, however, in that it is covalently linked to an inverse agonist, 11-*cis*-retinal, which ensures homogeneity of the inactive state. In response to light, 11-*cis*-retinal isomerizes to the agonist, all-*trans*-retinal, resulting in an allosteric transition that culminates in a 6–8 Å movement of transmembrane helix 6 (TM6) away from the transmembrane bundle, thereby creating an intracellular transducer-binding site. Other general structural changes that accompany GPCR activation include a contraction of the top of the receptor, an inward movement of TM5 and TM7, and a rotation of TM3[4]. These changes are likely to be facilitated by rearrangements in a conserved network of 'microswitches' (Box 1, top) and water molecules[5], common to most class A GPCRs, and are key components of allosteric communication between the orthosteric and intracellular transducer sites, which are typically located ~40 Å or more apart. A second unique feature of rhodopsin is that its active conformation is essentially fully attained even in the absence of co-bound G protein[6]. By contrast, most other known GPCR structures indicate that GPCRs require binding of a high efficacy agonist and an intracellular state-selective binding partner, such as a G protein, G-protein mimetic or arrestin to attain their fully active state. Notably, these structural observations are in agreement with pharmacological studies that were carried out nearly 40 years ago, which first identified a ternary complex between agonist, receptor and G protein as the molecular basis for drug efficacy at GPCRs[7]. Collectively, these studies highlight that signal transduction by GPCRs is an allosteric process that involves distal communication between conformationally linked sites.

GPCRs adopt multiple conformational states

High-resolution structures of GPCRs are snapshots of a dynamic system, and a two-state formalism is unlikely to accommodate their full conformational repertoire. In addition, currently solved structures represent the most stable conformations under specific experimental conditions; since GPCR stabilization is often achieved using a combination of detergents, fusion proteins, antibodies or mutations, this may limit the conformations explored by the receptor, as supported by biophysical studies[8,9]. These factors highlight an important caveat for interpreting structural studies and the need for complementary approaches. Accordingly, breakthroughs in NMR spectroscopy and molecular dynamics simulations have added to our understanding of

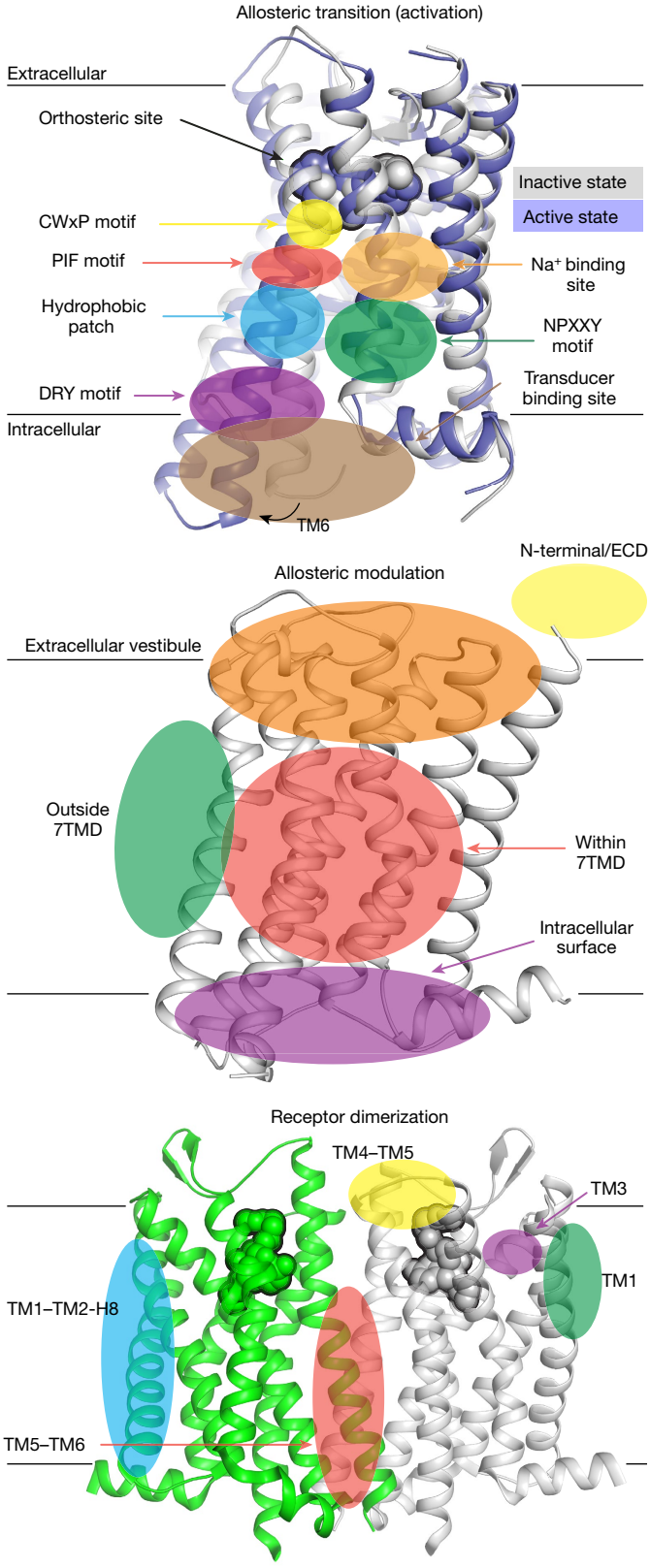
[1]Drug Discovery Biology, Monash Institute of Pharmaceutical Sciences, Monash University, Melbourne, Victoria, Australia. *e-mail: david.thal@monash.edu; arthur.christopoulos@monash.edu

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Box 1

The allosteric nature of GPCRs

There are three key facets to GPCR allostery. First, the change in conformational state of the receptor upon binding of orthosteric ligand or transducer to their respective sites is commonly defined as an 'allosteric transition'. Many studies, specifically of the rhodopsin-like class A GPCRs, have identified structural features associated with the allosteric transition, including a large outward movement of TM6 and a smaller inward movement of TM5 and TM7 mediated by a broadly conserved network of interactions that link the orthosteric and transducer binding sites. This allosteric network involves groups of neighbouring residues (microswitches); key examples include the CWXP motif, the PIF motif, an allosteric sodium-binding site, the NPXXY motif, and the E/DRY motif. The top panel in the figure shows these common features of the allosteric transition that activates class A GPCRs, mapped onto the inactive-state (white, PDB: 1F88) and active-state (blue, PDB: 3PQR) structures of rhodopsin. Second, because GPCRs accommodate a diverse array of orthosteric agonists (via different sites) within a common overall 7TMD architecture, indirect allosteric interactions or allosteric modulation between an orthosteric ligand and other conformationally-linked binding sites on the same receptor are possible because of the presence of multiple, functionally targetable domains. These allosteric modulatory sites can be grouped into five general categories, depending on the location of the orthosteric site. These allosteric modulatory sites are shown mapped onto the M_2 mAChR–iperoxo–LY2119620 structure (PDB: 4MQT) in the middle panel in the figure. Third, a common feature of the majority of allosteric proteins is an oligomeric architecture[1]. Although GPCRs represent a notable exception to this rule (with numerous examples of allostery occurring within a monomer), a body of experimental evidence highlights the potential for allosteric interactions between binding sites on GPCRs within dimeric or higher-order oligomeric arrays. Ovals in the bottom panel in the figure indicate key interfaces that have been implicated in this phenomenon in structural biology studies, mapped onto the dimeric structure of the μOR (PDB: 4DKL).



the conformational plasticity associated with GPCR signalling, with studies indicating the existence of multiple inactive- and active-state conformations[9–11]. Figure 1 summarizes some of these findings in terms of a 'continuum of conformations'.

Most GPCR structures solved to date are of inactive states. Unsurprisingly, the largest differences between the inactive-state structures are found in the extracellular regions, from the orthosteric pocket extending to the extracellular loops (ECLs), and are reflective of diversity in ligand binding across the GPCR superfamily. Otherwise, inactive-state structures are broadly similar, particularly with respect to intracellular regions. Of note, ^{19}F-NMR studies on the β_2 adrenergic receptor (β_2AR) and the adenosine A_{2A} receptor (A_{2A}R) describe two rapidly converting inactive conformations representing a stabilized and a broken 'ionic lock', respectively, between TM3 and TM6[10,11]. This lock

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

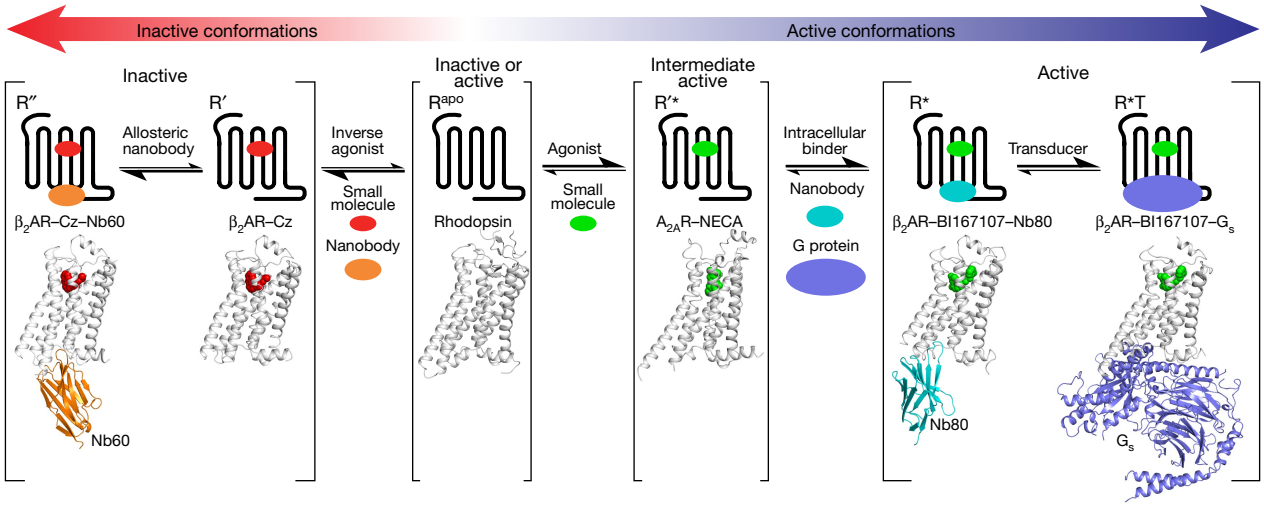


Fig. 1 | A continuum of GPCR conformational states. Representative examples of apo, ligand-, transducer- and/or nanobody (Nb)-bound GPCR structures, highlighting the potential diversity of conformational states. Inactive states (R″, R′) exemplified by β_2AR bound to the inverse agonist carazolol (Cz, red spheres; PDB: 2RH1) alone or together with the inverse agonist Nb60 (orange; PDB: 5JQH). For R^{apo}, the structure of rhodopsin in an active conformation that includes a molecule of detergent (not shown) in the orthosteric binding site is shown (PDB: 4J4Q). There are several structures of GPCRs in intermediate-active conformations (R′*), as exemplified by the A_{2A}R bound to the agonist 5-*N*-ethyl-carboxamidoadenosine (NECA) (green spheres; PDB: 2YDV). Structures of multiple GPCRs have also been determined in active-state complexes with either nanobodies (R*) or transducers (R*T), as exemplified by the β_2AR determined in complex with Nb80 (cyan, PDB: 3P0G) or the heterotrimeric G_s (blue, PDB: 3SN6).

can be further stabilized by the co-binding of inactive-state-specific nanobodies or antibodies[12,13], indicating that even so-called 'inactive states' can themselves display conformational heterogeneity. Although there are more than 50 agonist-bound structures, many of these more closely resemble an inactive rather than active conformation because they lack an additional binding partner, such as a G protein or arrestin, to allosterically stabilize the fully active conformation (R*). However, some of these agonist-bound states may represent intermediate conformations (R′*) between the transition from inactive (R′) to active (R*) conformations. General features of intermediate conformations include rearrangement of residues in the orthosteric site, microswitch regions, and possible partial outward movement of TM6. There are also multiple structures of agonist-bound GPCR–G-protein ternary complexes, and a structure of rhodopsin bound to arrestin (Supplementary Table 2). These structures unambiguously reveal the full extent of intracellular movement that is required for transducer binding.

By contrast, there are few structures of ligand-free GPCRs. Rhodopsin was the first to be reported, and multiple active-state opsin structures have also been determined, although a detergent molecule occupies the orthosteric site in several of these structures[14]. Several other receptors have had their structures determined in apparently ligand-free inactive conformations, however these finding remain controversial[15–17]. These studies highlight the challenges currently facing the field in obtaining true apo GPCR structures; however, such structures, if solved, could reveal novel GPCR conformations that facilitate 'state-selective' structure-based drug discovery that targets allosteric sites, rather than typical orthosteric-focused programs that (by default) require a ligand-bound structure. Nevertheless, the existing repertoire of active, inactive, intermediate and possibly apo GPCR structures supports the hypothesis that GPCRs adopt multiple conformational states that can be differentially stabilized by molecules binding to topographically disparate sites, either alone or in combination.

Exogenous allosteric modulators

One of the most exciting areas in modern drug discovery is the pursuit of molecules that bind to spatially distinct allosteric sites to modulate the actions of orthosteric ligands. Positive allosteric modulators (PAMs) enhance activity and negative allosteric modulators (NAMs) inhibit activity; whereas neutral allosteric ligands (NALs) have no net effect on the activity of orthosteric ligands, but competitively block the actions of PAMs or NAMs that bind to the same allosteric site[18]. Allosteric ligands may mediate agonism or inverse agonism in their own right, with or without PAM, NAM or NAL properties, suggesting novel pharmacological opportunities for drug development[1] (Box 2). 'Bitopic' molecules have also been described, which bridge both orthosteric and allosteric sites on a single GPCR; such an interaction was structurally validated by a recent bitopic ligand-bound structure of the leukotriene B4 receptor[19]. Allosteric drug discovery targeting GPCRs has largely been driven by classic pharmacological or biochemical approaches, but the last five years have witnessed an explosion in structural understanding of the phenomenon (Fig. 2, Table 1). Although a general classification of GPCR allosteric sites can be complicated because such sites must, by necessity, be considered in relation to the widely divergent orthosteric sites between GPCRs, it is clear that druggable allosteric sites span the entire receptor surface, encompassing at least five key regions (Fig. 2, Box 1, middle):

1. The N-terminal ECD

This region is particularly relevant to the ECD of class B, C and F GPCRs. To our knowledge, there are currently no solved structures with synthetic modulators in this region, but there are structures bound to endogenous allosteric modulators (discussed below).

2. The extracellular vestibule

This region encompasses the ECLs and the classic (class A) biogenic amine orthosteric site in the 7TMD. A prototypical example is the M_2 muscarinic acetylcholine receptor[20] (mAChR; Box 1, middle; Fig. 2). This structure highlights a mechanism whereby activation promotes a closure of the extracellular vestibule, which is preferentially stabilized by PAMs, a contraction of the orthosteric site, and the opening of the intracellular region for transducer binding. Recent studies on the β_2AR suggest that G-protein binding to the inside of the GPCR allosterically stabilizes a similar closed conformation at the extracellular end of the receptor[21]. Therefore, stabilization of a closed extracellular vestibule may be a common mechanism by which PAMs can increase agonist activity via this region. Conversely, molecular dynamics simulations of inactive-state M_2 mAChRs suggest that NAMs binding in the vestibule have the opposite effect; these tend to be larger molecules that favour a more open vestibule[22]. A similar NAM-bound structure has been reported for proteinase-activated receptor 2 (PAR2), although there

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Box 2

Therapeutic applications of allosteric modulators

There are numerous advantages of allosteric modulators as therapeutic agents in comparison to classic orthosteric agonists or antagonists[1]. The first is the potential for on-target selectivity; that is, a higher degree of drug selectivity between subtypes of a given GPCR, owing to a lower degree of amino acid conservation within the allosteric site relative to the orthosteric site, and/or selective cooperativity (PAM or NAM effect) exerted at one GPCR subtype but not others. Second, the effects of allosteric drugs may reach a saturable limit, which is governed by the magnitude of cooperativity between orthosteric and allosteric sites. This provides an opportunity to fine tune the actions of an allosteric medicine to better tailor the drug to both its target GPCR and disease indication, with reduced propensity for dose-related on-target side effects. Third, 'pure' PAMs or NAMs remain quiescent in the absence of tonic endogenous agonist(s), and only exert their effects when and where neurotransmitters or hormones are released. This has the twofold advantage of exploiting homeostatic spatio-temporal patterns associated with physiological signalling, as well as providing an avenue for tissue-specific effects as a result of the reciprocal nature of allosteric interactions—the higher the endogenous hormonal levels are in a given region, the more sensitive the receptor will be to the allosteric drug. A limitation of these latter mechanisms is in diseases where endogenous agonist tone is markedly depleted (for example, late stage neurodegeneration) to the point that it can no longer be modulated. However, allosteric agonists may still prove useful in such circumstances by virtue of their greater on-target selectivity. Finally, allosteric ligands have the potential to impose biased signalling on the actions of endogenous agonists. Although many of these theoretical advantages have yet to be validated clinically, a number of GPCR modulators are present at different stages of development[1]. Importantly, the recent expansion in structural knowledge of GPCR allosteric sites is also being leveraged through rational drug-design studies[104], providing a proof-of-concept that selective modulators can be discovered using structure-based approaches on fragment[29] or small molecule libraries[105]. This includes the provocative potential for combining existing potent, yet non-selective, orthosteric medicines with allosteric drugs so that they may be repurposed[106].

is a lack of structural and pharmacological comparisons with its agonist[23] (Fig. 2, Table 1). In addition, the HIV drug maraviroc has been described as a NAM for the chemokine receptor CCR5 because it binds deeper into the 7TMD, away from the primary extracellular chemokine recognition site[24]. However, recent structure-function studies and a CCR5–chemokine antagonist structure suggest that this site may overlap with part of the orthosteric site, bringing into question claims of an allosteric mode of action of maraviroc[25,26].

3. Within the 7TMD

The best example of allosteric modulators binding deeper within the 7TMD is from 7TMD structures of the class C metabotropic glutamate receptors $mGlu_1$[27] and $mGlu_5$[28] in complex with the NAMs FITM and mavoglurant, respectively, and further validated by newer structures[29,30]. These NAMs stabilize an inactive state through a cavity within the 7TMD (Fig. 2, Table 1), which may generically possess more than one allosteric site, as indicated by the structure of the class F Smoothened receptor (SMO). The NAM, SANT-1, adopts a similar pose to that seen in the $mGlu_1$–FITM structure[31]. The deepest allosteric pocket within the 7TMD identified thus far is exemplified by CP-376395, a NAM of the class B corticotropin-releasing factor 1 receptor (CRF_1R), which acts by sterically preventing formation of a kink in the conserved $Pro^{6.47b}$-X-X-$Gly^{6.50b}$ (superscript refers to the Wootten numbering system for class B GPCRs[32]) motif in TM6, which is vital for the activation of class B GPCRs[33].

4. Outside the 7TMD extending to the bilayer

Several extrahelical allosteric sites have been structurally confirmed (Table 1). A common pocket was identified outside TM5–7 near the intracellular surface for NAMs that target the glucagon receptor (GCGR)[34] and glucagon-like peptide-1 receptor (GLP-1R)[35,36] (Fig. 2). These NAMs prevent the outward movement of transmembrane helices that are required for activation. BPTU, a NAM of the purinergic receptor $P2Y_1$, binds in a shallow pocket defined by residues in TM1–3 and ECL1[37]. Similarly, AZ3451, a NAM of the PAR2 receptor, binds in a pocket created by residues in TM2–4[23]. Despite binding to different GPCRs, both NAMs are likely to share a common mechanism that prevents rotation and translation of TM3 (which is usually required for activation of class A GPCRs). Multiple allosteric agonist co-structures have been reported for the free fatty acid receptor GPR40, including with the drugs TAK-875[38] and MK-8666[39], which bind within the 7TMD but extend into the lipid bilayer through TM3–4, and a co-structure with both MK-8666 and the PAM AP8; AP8 binds in a second allosteric pocket between TM3 and TM5 above ICL2[39]. Recently, a structure of the complement C5a receptor bound to the NAM NDT9513727 was determined, revealing an allosteric site that closely resembles the GPR40–AP8 PAM site (Fig. 2). NDT9513727 is likely to function as a NAM because it acts as a steric 'wedge', preventing the allosteric transition of TM5 that is required for activation[40].

5. The intracellular surface

Structures of the chemokine receptors CCR2 and CCR9 in complex with their respective NAMs, CCR2-RA-[*R*][41] and vercirnon[42] revealed a druggable intracellular allosteric site below the middle of the 7TMD that spatially overlaps with the G-protein-binding site. These NAMs probably stabilize an inactive conformation of TM6 while sterically impeding receptor interactions with a G protein (and possibly arrestin). Most recently, a structure of the β_2AR in complex with a novel beta-blocker NAM, Cmpd-15-PA, also confirmed the existence of an intracellular allosteric site[43]. Of note, CCR2-RA-[*R*], vercirnon and Cmpd-15-PA all bind in very similar locations on the respective receptors, albeit with distinctly different chemical structures and interactions (Fig. 2, Table 1). The common binding site may therefore encompass a conserved intracellular allosteric site for small drug-like molecules. Although speculative, it may be possible to chemically modify such molecules to also promote PAM activity, although this may necessitate molecules that are better at opening this intracellular vestibule, which is the opposite mechanism to that of PAMs, which close the extracellular vestibule. An alternative mechanism has been proposed for the actions of the GLP-1R PAM BETP, which modifies a cysteine residue at the ICL3–TM6 junction to facilitate the activating outward movement of TM6[44].

Endogenous allosteric modulators

Given the diverse nature of GPCR binding pockets, it is possible that some of these may serve as sites for endogenous allosteric ligands. There is certainly pharmacological and biochemical evidence of such a role for various ions, amino acids, peptides, proteins, lipids and autoantibodies in both health and disease[45], but a general mechanistic understanding of how these putative modulators alter GPCR function is largely lacking and beyond the scope of this Review. Instead, we highlight three specific examples of endogenous ligands that have benefited from structural breakthroughs. The first is in the general field of 'accessory proteins', which we broadly define as endogenous proteins that interact with specific GPCRs to regulate their binding, signalling or trafficking properties in a ligand-, pathway- or cell-specific manner. One of the best examples is a family of three single membrane-spanning proteins known as 'receptor activity-modifying proteins' (RAMPs),

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

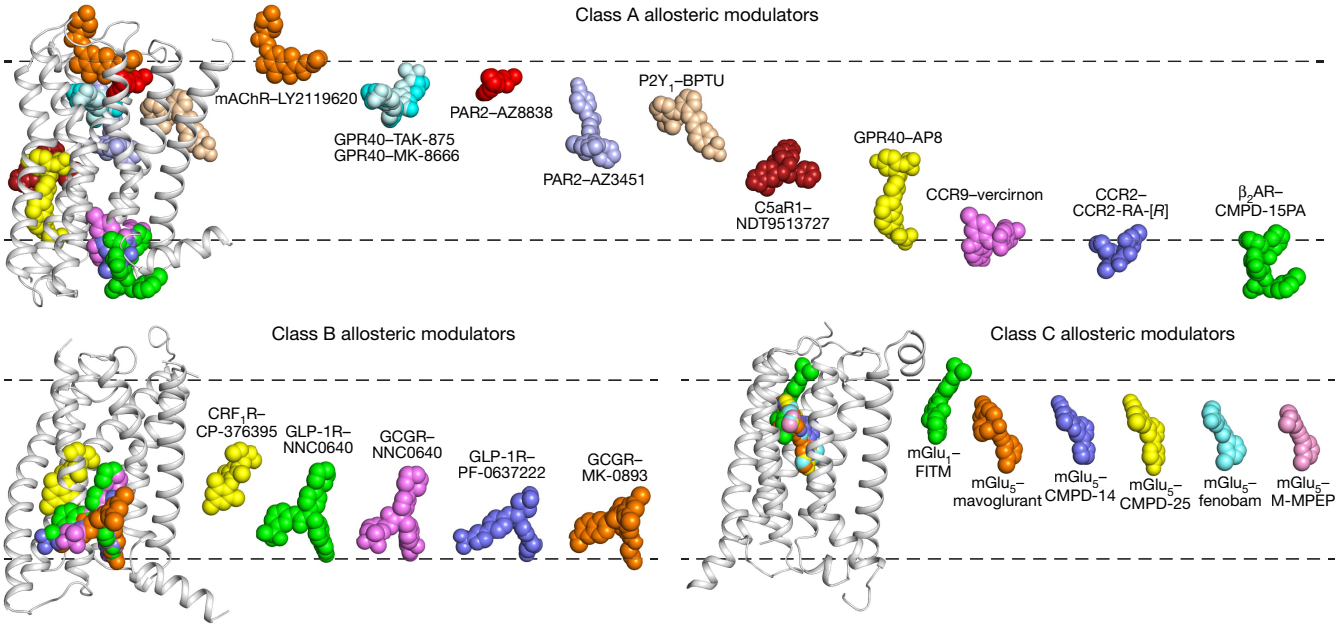


Fig. 2 | Diversity in the binding sites of synthetic allosteric modulators across GPCR classes. Chemical structures of allosteric modulators (coloured spheres) mapped onto representative family members of class A (M_2 mAChR, PDB: 4MQT), class B (GCGR, PDB: 5EE7) and class C ($mGlu_5$, PDB: 4OO9) GPCRs. Dashed lines indicate the boundary of the lipid bilayer. Further information is listed in Table 1.

which associate with various GPCRs to modulate their physiological and pharmacological phenotype[46]. Although no structures of RAMPs in complex with full-length GPCRs have been solved, there are structures of RAMP1 or RAMP2 in complex with the ECD of the calcitonin receptor-like receptor, which unambiguously reveal a RAMP–GPCR interface that forms a drug binding site[46]. A second important example of endogenous allosteric modulators is that of ions. Indeed, sodium (Na^+) was shown to negatively modulate agonist binding at opioid receptors more than 40 years ago[47], and its allosteric effects have since been observed with other class A GPCRs[45,48]. A recent high-resolution structure of the $A_{2A}R$[49] revealed an intramembranous Na^+ binding site and associated water cluster. Similar Na^+ sites have been observed in other GPCRs (Supplementary Table 3). The functional role of the intramembranous Na^+ site is unclear, although it is possible that Na^+ may be acting as a cofactor, stabilizing the ligand-free state of some GPCRs and balancing the active-state–inactive-state allosteric transition[50]. Importantly, the drug amiloride and its derivatives bind to the Na^+ site, albeit with low affinity, suggesting that this site may be targetable by drugs[19,48].

Finally, there is a growing interest in the role of lipids in GPCR function, given that they are membrane proteins. To date, most structural studies have focused on the role of cholesterol as a putative allosteric modulator of GPCRs. Biochemically, cholesterol increases the thermal stability of rhodopsin and the β_2AR[51], and the binding affinities

Table 1 | Reported structures of GPCRs in complex with synthetic allosteric modulators

GPCR	Modulator	Site of interaction	PDB code	Notes	Ref.
Class A					
β_2AR	CMPD-15PA	Intracellular surface	5X7D	Co-bound with carazolol, intracelluar surface near TM6/TM7/H8	43
C5aR1	NDT9513727	Outside the 7TMD	5O9H	Wedged in dimeric interface between TM3/TM4/TM5	40
CCR2	CCR2-RA-[*R*]	Intracellular surface	5T1A	Co-bound with BMS-681, intracelluar surface near TM6/TM7/H8	41
CCR9	vercirnon	Intracellular surface	5LWE	Intracellular surface near TM6/TM7/H8	42
GPR40	TAK-875	Outside the 7TMD	4PHU	Wedged between TM3/TM4 and extending into bilayer	38
GPR40	MK-866	Outside the 7TMD	5TZR	Wedged between TM3/TM4 and extending into bilayer	39
GPR40	AP8	Outside the 7TMD	5TZY	Co-bound with MK-8666, outside 7TMD near TM3/TM4/TM5	39
M2 mAChR	LY2119620	Extracellular vestibule	4MQT	Co-bound with iperoxo, top of extracellular vestibule	20
$P2Y_1$	BPTU	Outside the 7TMD	4XNV	Outside 7TMD near TM1/TM2/TM3	37
PAR2	AZ8838	Extracellular vestibule	5NDD	Floor of vestibule between ECL1/2 near TM1/TM2/TM3/TM7	23
PAR2	AZ3451	Outside the 7TMD	5NDZ	Outside 7TMD near TM2/TM3/TM4	23
Class B					
CRF_1R	CP-376395	Within the 7TMD	4K5Y	Deep within 7TMD wedged between TM3/TM5/TM6	33
GCGR	NNC0640	Outside the 7TMD	5XEZ, 5XF1	Near intracellular surface, outside 7TMD near TM6/TM7	35
GCGR	MK-0893	Outside the 7TMD	5EE7	Near intracellular surface, outside 7TMD near TM6/TM7	34
GLP-1R	NNC0640	Outside the 7TMD	5VEX	Near intracellular surface, outside 7TMD near TM6/TM7	36
GLP-1R	PF-0637222	Outside the 7TMD	5VEW	Near intracellular surface, outside 7TMD near TM6/TM7	36
Class C					
$mGlu_1$	FITM	Within the 7TMD	4OR2	Middle of 7TMD	27
$mGlu_5$	mavoglurant	Within the 7TMD	4OO9	Middle of 7TMD	28
$mGlu_5$	CMPD-25	Within the 7TMD	5CGC	Middle of 7TMD	29
$mGlu_5$	CMPD-14	Within the 7TMD	5CGD	Middle of 7TMD	29
$mGlu_5$	fenobam	Within the 7TMD	6FFH	Middle of 7TMD	30
$mGlu_5$	M-MPEP	Within the 7TMD	6FFI	Middle of 7TMD	30
Class F					
SMO	SANT-1	Within the 7TMD	4N4W	Middle of 7TMD	31

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

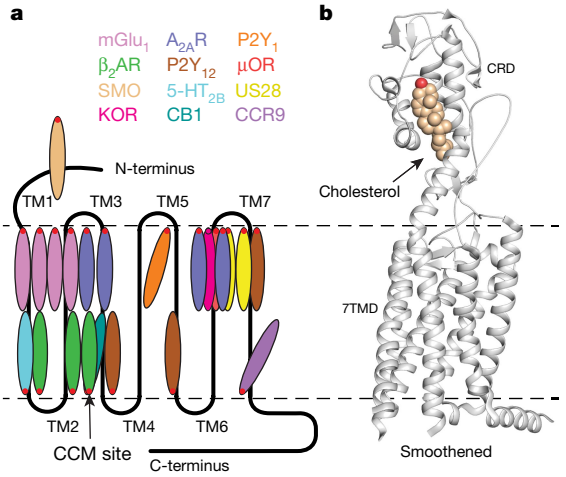


Fig. 3 | Cholesterol-binding sites across GPCRs. Cholesterol molecules have been observed bound to GPCRs in more than 40 PDB entries for 12 different receptors (Supplementary Table 2). **a**, Schematic of cholesterol-binding sites shown as ovals, coloured according to the GPCR subtype, with the red dot indicating the hydroxyl group of cholesterol. The arrow indicates a proposed cholesterol consensus motif (CCM). Positions of cholesterol are mapped onto their approximate location on the 7TMD cartoon. **b**, Cholesterol is an endogenous allosteric modulator of Smoothened (PDB: 5L7D), and binds to the cysteine-rich domain (CRD). Dashed lines indicate the boundary of the lipid bilayer.

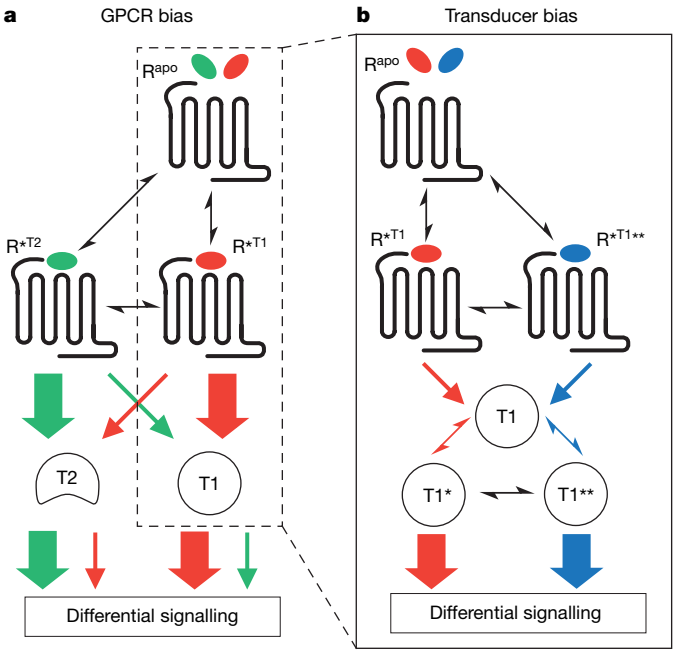


Fig. 4 | Conformational mechanisms of biased agonism. a, Biased agonism as a consequence of conformational selection by ligand occurring at the level of GPCR conformation. Two different agonists (red and green) preferentially select different active states (R^{*T1} and R^{*T2}), which are characterized by different strengths of coupling (indicated by arrow size) to transducer proteins T1 and T2. **b**, Biased agonism as a consequence of conformational selection occurring at the level of the transducer protein. Two different agonists (red and blue) promote a similar active state (R*) with a given preference for transducer T1, but bias arises through allosteric communication between the receptor, ligand and transducer as the latter also undergoes conformational selection (T1*, T1**).

of some agonists at other GPCRs[52]. Crystal structures of the β_2AR reveal three cholesterol binding sites; one of these, located in a cleft between TM2 and TM4, has been termed a 'cholesterol consensus motif' (CCM) owing to the high sequence prevalence (44%) across class A GPCRs[51]. Molecular dynamics simulations suggest that cholesterol allosterically modulates the β_2AR by limiting its conformational flexibility[53]; this is supported by the necessity for cholesterol for crystallization of the receptor[51,54]. Indeed, cholesterol is present in over 40 PDB entries, occupying over 20 distinct sites (Fig. 3a, Supplementary Table 3). However, the only other structures with cholesterol bound near the region proposed to contain the CCM are those of the cannabinoid receptor CB_1[55] and the purinergic receptor $P2Y_{12}$[56], both of which lack an actual CCM. Therefore, at this stage, there is a lack of crystallographic evidence for generalization of cholesterol binding to distinct sites across all GPCRs, other than to note that cholesterol is attracted to transmembrane grooves and hydrophobic patches[52]. By contrast, several reported structures show cholesterol binding to the extracellular cysteine-rich domain (CRD) of SMO[57,58], which allosterically activates the 7TMD and mediates downstream signalling[57,59] (Fig. 3b). It is therefore likely that the potential role and druggability of cholesterol-binding sites across various GPCRs will remain a fertile area of research in future. Irrespective of the type of endogenous substance, it should be noted that if endogenous allosteric modulators can be proven to mediate disease, then entirely new classes of medicines based on NALs may be pursued in such cases to specifically target the endogenous allosteric ligand site while sparing the actions of the cognate orthosteric agonist[45]. This is because NALs will act as competitive antagonists of endogenous disease-mediating substances that bind to the same allosteric site, while having no effect on normal signalling mediated by the orthosteric agonist.

Biological agents as allosteric modulators

One of the major methodological breakthroughs for ensuring conformational homogeneity in GPCR structural biology has been the development of biological agents, such as antibodies, antibody fragments (Fabs) and nanobodies, to stabilize distinct active or inactive states in the presence of co-bound ligand[60]. Such biologic tools are intrinsically allosteric. For example, nanobody 80 (Nb80), which facilitated the solution of the first active-state β_2AR structure (Supplementary Table 2) binds to the intracellular surface of the β_2AR in a manner analogous to the C-terminal α-helix of G_s and acts as a PAM of agonist binding. Allosteric nanobodies have facilitated the crystallization of active conformations of other GPCRs, including the M_2 mAChR, the μ-opiod receptor (μOR), viral chemokine receptor US28 and the κ-opioid receptor (Supplementary Table 2). Biological agents have also been used to stabilize heterotrimeric G proteins[61] and β-arrestins[62]. This breakthrough enabled the first structure of a GPCR–G protein complex, and has since been extended to assist in solution of the first cryo-EM structures of GPCR–G protein complexes[63,64]. Although it may be argued that biological agents that recognize intracellular epitopes or proteins may not be directly translatable as therapeutics, their utility is not solely limited to structural studies[60]. For example, lipidated peptide fragments from GPCRs, termed 'pepducins', interact as intracellular allosteric modulators and are currently in clinical trials[65]. Moreover, newer biological agents have emerged as allosteric modulators, such as RNA aptamers that stabilize ligand-specific conformations[66]. Finally, the potential for extracellular allosteric biological agents should not be overlooked as a promising new area of research, as evidenced by the ongoing and successful clinical development of therapeutic antibodies[67].

Biased agonism: cytosol-directed allostery

Key aspects of the allosteric nature of GPCRs are required for both a ligand and transducer to ensure stimulus transfer and reciprocity in communication between ligand and transducer sites. Although these mechanisms have been appreciated for decades[7,68–70], diversity of cellular outcomes imparted by drugs (for example, full versus partial agonists) acting at the same GPCR was traditionally ascribed to post-receptor mechanisms; however, this concept is outdated. Following activation, GPCRs are phosphorylated by GPCR kinases (GRKs) in a GPCR-region-specific manner, leading to recruitment

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

of arrestins to terminate G-protein signalling and initiate alternative waves of signalling[71]. Convergent data indicating that GPCRs adopt multiple ligand-stabilized active states, each linked to distinct physiological outcomes through promiscuity in their preferential choice of transducer to the relative exclusion of others, have led to the phenomenon termed 'biased agonism' or 'functional selectivity'[71], and is transforming drug discovery. However, these phenomena are natural consequences of allostery. The mechanistic trigger underlying biased agonism is the tripartite communication between ligand, GPCR and transducer; this link has been explicitly referred to as 'biased allostery'[72] (Fig. 4a). It is therefore timely to consider the extent to which structural studies have addressed the molecular basis of GPCR-biased agonism.

Much research on GPCR–transducer interactions has focused on G proteins, although much remains to be learned with regards to the molecular determinants of G protein-mediated efficacy and selectivity[73]. By contrast, there is a relative lack of detailed structural information regarding direct interactions between GPCRs and GRKs, despite the central role that GRKs play in orchestrating interactions between GPCRs, G proteins and arrestin[74]; recent studies suggest the involvement of the 'regulator of G-protein signalling homology domain' of GRKs as a recognition site for GPCRs[75,76]. The interaction of arrestins with the GPCR is typically described by a two-step process[77,78] involving binding of the C-tail of arrestin with the phosphorylated C terminus of the GPCR (the 'hanging conformation'[62,79,80]), followed by rotation of the N- and C-terminal domains of arrestin to engage with the intracellular core of the GPCR (the 'core conformation')[80,81]. These events have significant implications for arrestin-mediated biased agonism. First, the strength of the initial interaction may be differentially encoded in a ligand-dependent manner via a 'phosphorylation barcode' on the GPCR[81,82]. Second, although binding of arrestin and G-protein binding to GPCRs have been considered to be mutually exclusive, the hanging conformation of arrestin may allow a GPCR to simultaneously engage with arrestin and a G protein, as suggested by a recently discovered macromolecular 'megaplex' with unique signalling properties[83]. Finally, the fully engaged arrestin–receptor core conformation, which is not compatible with G-protein binding, can nevertheless participate in β-arrestin-specific biased agonism.

A complete structural understanding of biased agonism requires the ability to decipher molecular interactions formed between biased ligands, GPCR and transducer. Although our understanding of how this occurs is in its infancy, some advances have been made. For example, structures of the 5-hydroxytryptamine receptors 5-HT_{1B} and 5-HT_{2B} bound to ergotamine (ERG) have been determined; the 5-HT_{1B} complex is weakly biased towards β-arrestin, whereas the 5-HT_{2B} complex is strongly biased towards β-arrestin[84,85], relative to the complex with the endogenous agonist 5-hydroxytryptamine. The structure of 5-HT_{2B} in complex with LSD, another strongly β-arrestin biased ligand, has also been determined[86]. The 5-HT_{2B} structures share common features that are not observed in the 5-HT_{1B} structure. Differential ligand contacts[84] may explain the different GPCR conformations. Differential ligand-receptor contacts were also proposed to underlie the mechanism of the biased ligand carvedilol in complex with β_1AR[87].

Perhaps the most important conclusion of such structural studies to date is that they provide insights into the most proximal triggers of biased agonism. However, a challenge for the future is the need to incorporate the role of the transducer, which remains a non-trivial task. Nonetheless, recent intramolecular FRET and BRET studies focusing on arrestin loop movements[88,89] have suggested that arrestin adopts multiple states upon receptor binding. Moreover, different agonists acting on the same GPCR produce distinct signatures, indicating that arrestin conformations can be influenced by GPCR ligands. Since activated arrestin conformations persist for some time after dissociation from receptors[88], this may lead to differential engagement of downstream effectors, resulting in biased agonism, consistent with recent studies highlighting a role of kinetics in the phenomenon[90,91]. Perhaps more provocatively, similar arguments can be made for G proteins beyond their role as mediators of nucleotide exchange in a manner that is dictated solely by their affinity for a given active state. For example, salmon and human calcitonin promote distinct conformations of $G\alpha_s$ in complex with the calcitonin GPCR, each with differential GTP sensitivities, distinct receptor–G protein complex lifetimes and consequent differences in rates of cAMP formation[92]. Likewise, a comprehensive study of different β_2AR ligands found that they can differentially affect movement of TM6, engagement by $G\alpha_s$, and differential efficiencies of nucleotide exchange[93]. Furthermore, a recent cryo-EM structure of GLP-1R in complex with the biased peptide agonist, exendin-P5, revealed a different ECL3 conformation along with a 6° difference in the angle of G-protein engagement with the receptor[94]. Together, these findings suggest that to fully appreciate the molecular nature of GPCR efficacy and biased agonism, it is necessary to extend the concept of conformational selection by ligands from the level of ligand–GPCR to incorporate selection of transducer conformations through allosteric interactions (Fig. 4b).

GPCR dimerization

Although the formation of dimers and higher-order oligomers is well established as a universal paradigm mediating binding, signalling and allostery in other receptor superfamilies[1], the role of oligomerization of most GPCRs is not as clearly defined, as it is known that many have the potential to mediate signalling as monomers[95,96]. However, GPCR dimerization remains an intensively studied phenomenon and has been proposed to modulate trafficking, ligand binding cooperativity, and signalling efficacy[97]. Of particular relevance to the current Review, the degree to which structural studies unambiguously support GPCR dimerization as a molecular mechanism underlying ligand recognition, cooperativity, or allosteric transitions tends to vary with the receptor subfamily. For instance, multiple parallel dimeric interfaces have been observed in class A GPCR structures[97] (Box 1, bottom). The different interfaces identified thus far in many of these structures may reflect artefacts of the crystallization process, and caution should therefore be used when linking these observations to functional studies. For class B GPCRs, a functional role for dimerization, particularly involving TM4, has been proposed from mutagenesis, resonance energy transfer and pharmacological experiments[98].

Structurally, it is perhaps most instructive to consider the class C GPCRs, such as the metabotropic glutamate receptors (mGluRs), type B γ-aminobutyric acid receptor ($GABA_B$), Ca^{2+}-sensing receptor (CaSR) and taste receptors (T1Rs), because these receptors unambiguously function as obligate dimers[99]. Although there are currently no full-length class C GPCR structures, various inactive and active conformations of isolated venus flytrap domains from mGluRs, $GABA_B$, CaSR[100,101] and T1Rs[102] have been solved[99]. It is particularly noteworthy that these obligate dimeric GPCRs are all key environmental nutrient sensors that tightly regulate the activity of their cognate agonists. Thus, by extension, one may speculate that a key role for dimerization of GPCRs (even those known to function as monomers) is to constrain or bias the range of possible receptor activity in either a location-, tone- and/or context-dependent manner when required.

Outlook

Despite rapid recent advances in structural biology of GPCRs, there remains a great deal to be discovered to facilitate fundamental understanding of the role of allostery, and to fully realize the potential for allosteric and biased medicines as new therapeutic classes. For instance, a mechanistic understanding of the allosteric transition leading to differential G-protein and/or arrestin recruitment (and thus pathway-biased signalling or modulation) requires structures of multiple GPCR–transducer–effector complexes—ideally with minimal perturbations associated with the process of protein engineering or crystallization artefacts. This principle extends to the solution of full-length structures of larger GPCRs (for example, class C GPCRs) and associated complexes in order to better understand dimerization. The solution of multiple structures of the same receptor bound to different ligand classes and/or transducers remains largely elusive yet can prove enormously beneficial for structure-based design. Developments in

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

alternative approaches, particularly NMR, computational biology and single-molecule fluorescence, as well as ongoing rigorous application of mutagenesis, chemical biology and analytical pharmacology, are likely to make substantial contributions to understanding the rich repertoire of GPCR allosteric effects. Structural validation of the enormous diversity of allosteric modulator-binding sites, as well as common motifs for such sites among disparate GPCRs, also raises two important questions for the field. First, what are the mechanistic drivers of the selectivity associated with many allosteric modulators if the sites are more prevalent or conserved than previously appreciated? Second, is there a role for endogenous modulators in health and disease, and how can this be exploited? Encouragingly, it is worth appreciating that many of the mechanisms associated with allostery of GPCRs are likely to be universal, and thus have either been better studied in some respects, or remain to be exploited in others, in different receptor superfamilies[1,103]. This suggests that there is very fertile ground for cross-disciplinary breakthroughs in understanding the ongoing role of allostery in the life sciences.

Received: 1 December 2017; Accepted: 17 May 2018;
Published online 4 July 2018.

1. Changeux, J. P. & Christopoulos, A. Allosteric modulation as a unifying mechanism for receptor function and regulation. *Cell* **166**, 1084–1102 (2016).
This review highlights the key principles of allostery as common elements across multiple receptor superfamilies
2. Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schioth, H. B. & Gloriam, D. E. Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discov.* **16**, 829–842 (2017).
3. Tesmer, J. J. Hitchhiking on the heptahelical highway: structure and function of 7TM receptor complexes. *Nat. Rev. Mol. Cell Biol.* **17**, 439–450 (2016).
4. Tehan, B. G., Bortolato, A., Blaney, F. E., Weir, M. P. & Mason, J. S. Unifying family A GPCR theories of activation. *Pharmacol. Ther.* **143**, 51–60 (2014).
5. Yuan, S., Filipek, S., Palczewski, K. & Vogel, H. Activation of G-protein-coupled receptors correlates with the formation of a continuous internal water pathway. *Nat. Commun.* **5**, 4733 (2014).
6. Choe, H. W. et al. Crystal structure of metarhodopsin II. *Nature* **471**, 651–655 (2011).
7. Kent, R. S., De Lean, A. & Lefkowitz, R. J. A quantitative analysis of beta-adrenergic receptor interactions: resolution of high and low affinity states of the receptor by computer modeling of ligand binding data. *Mol. Pharmacol.* **17**, 14–23 (1980).
8. Eddy, M. T., Didenko, T., Stevens, R. C. & Wuthrich, K. β_2-adrenergic receptor conformational response to fusion protein in the third intracellular loop. *Structure* **24**, 2190–2197 (2016).
9. Van Eps, N. et al. Conformational equilibria of light-activated rhodopsin in nanodiscs. *Proc. Natl Acad. Sci. USA* **114**, E3268–E3275 (2017).
10. Manglik, A. et al. Structural insights into the dynamic process of β_2-adrenergic receptor signaling. *Cell* **161**, 1101–1111 (2015).
11. Ye, L., Van Eps, N., Zimmer, M., Ernst, O. P. & Prosser, R. S. Activation of the A_{2A} adenosine G-protein-coupled receptor by conformational selection. *Nature* **533**, 265–268 (2016).
These NMR studies on rhodopsin[9], β_2AR[10] and A_{2A}R[11] demonstrate the highly dynamic nature of the GPCR conformational landscape
12. Staus, D. P. et al. Allosteric nanobodies reveal the dynamic range and diverse mechanisms of G-protein-coupled receptor activation. *Nature* **535**, 448–452 (2016).
This study reports two distinct inactive state conformations and provides direct structural evidence for a dynamic range of both active and inactive GPCR states
13. Hino, T. et al. G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature* **482**, 237–240 (2012).
14. Carpenter, B. & Tate, C. G. Active state structures of G protein-coupled receptors highlight the similarities and differences in the G protein and arrestin coupling interfaces. *Curr. Opin. Struct. Biol.* **45**, 124–132 (2017).
15. Shihoya, W. et al. Activation mechanism of endothelin ETB receptor by endothelin-1. *Nature* **537**, 363–368 (2016).
16. Huang, J., Chen, S., Zhang, J. J. & Huang, X. Y. Crystal structure of oligomeric β_1-adrenergic G protein-coupled receptors in ligand-free basal state. *Nat. Struct. Mol. Biol.* **20**, 419–425 (2013).
17. Leslie, A. G., Warne, T. & Tate, C. G. Ligand occupancy in crystal structure of β_1-adrenergic G protein-coupled receptor. *Nat. Struct. Mol. Biol.* **22**, 941–942 (2015).
18. Christopoulos, A. et al. International Union of Basic and Clinical Pharmacology. XC. Multisite pharmacology: recommendations for the nomenclature of receptor allosterism and allosteric ligands. *Pharmacol. Rev.* **66**, 918–947 (2014).
This document provides the official nomenclature for ligand classification
19. Hori, T. et al. Na^+-mimicking ligands stabilize the inactive state of leukotriene B_4 receptor BLT1. *Nat. Chem. Biol.* **14**, 262–269 (2018).
20. Kruse, A. C. et al. Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **504**, 101–106 (2013).
This landmark study reports the first crystal structure of an active-state agonist-bound GPCR in complex with a synthetic allosteric modulator
21. DeVree, B. T. et al. Allosteric coupling from G protein to the agonist-binding pocket in GPCRs. *Nature* **535**, 182–186 (2016).
This paper provides direct structural evidence of the reciprocal allosteric coupling between G-protein and agonist-binding sites on a GPCR
22. Dror, R. O. et al. Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. *Nature* **503**, 295–299 (2013).
This key computational biology study highlights the utility of molecular dynamics for identifying mechanisms underlying the actions of allosteric modulation at a GPCR
23. Cheng, R. K. Y. et al. Structural insight into allosteric modulation of protease-activated receptor 2. *Nature* **545**, 112–115 (2017).
24. Tan, Q. et al. Structure of the CCR5 chemokine receptor–HIV entry inhibitor maraviroc complex. *Science* **341**, 1387 (2013).
25. Huma, Z. E. et al. Key determinants of selective binding and activation by the monocyte chemoattractant proteins at the chemokine receptor CCR2. *Sci. Signal.* **10**, eaai8529 (2017).
26. Zheng, Y. et al. Structure of CC chemokine receptor 5 with a potent chemokine antagonist reveals mechanisms of chemokine recognition and molecular mimicry by HIV. *Immunity* **46**, 1005–1017.e1005 (2017).
27. Wu, H. et al. Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator. *Science* **344**, 58 (2014).
28. Dore, A. S. et al. Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. *Nature* **511**, 557–562 (2014).
29. Christopher, J. A. et al. Fragment and structure-based drug discovery for a class C GPCR: Discovery of the mGlu5 negative allosteric modulator HTL14242 (3-chloro-5-[6-(5-fluoropyridin-2-yl)pyrimidin-4-yl]benzonitrile). *J. Med. Chem.* **58**, 6653–6664 (2015).
30. Christopher, J. A. et al. Structure-based optimization strategies for G protein-coupled receptor (GPCR) allosteric modulators: a case study from analyses of new metabotropic glutamate receptor 5 ($mGlu_5$) X-ray structures. *J. Med. Chem.* (2018).
31. Wang, C. et al. Structural basis for Smoothened receptor modulation and chemoresistance to anticancer drugs. *Nat. Commun.* **5**, 4355 (2014).
32. Wootten, D. et al. Polar transmembrane interactions drive formation of ligand-specific and signal pathway-biased family B G protein-coupled receptor conformations. *Proc. Natl. Acad. Sci. USA* **110**, 5211–5216 (2013).
33. Hollenstein, K. et al. Structure of class B GPCR corticotropin-releasing factor receptor 1. *Nature* **499**, 438–443 (2013).
34. Jazayeri, A. et al. Extra-helical binding site of a glucagon receptor antagonist. *Nature* **533**, 274–277 (2016).
35. Zhang, H. et al. Structure of the full-length glucagon class B G-protein-coupled receptor. *Nature* **546**, 259–264 (2017).
36. Song, G. et al. Human GLP-1 receptor transmembrane domain structure in complex with allosteric modulators. *Nature* **546**, 312–315 (2017).
37. Zhang, D. et al. Two disparate ligand-binding sites in the human P2Y1 receptor. *Nature* **520**, 317–321 (2015).
38. Srivastava, A. et al. High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875. *Nature* **513**, 124–127 (2014).
39. Lu, J. et al. Structural basis for the cooperative allosteric activation of the free fatty acid receptor GPR40. *Nat. Struct. Mol. Biol.* **24**, 570–577 (2017).
40. Robertson, N. et al. Structure of the complement C5a receptor bound to the extra-helical antagonist NDT9513727. *Nature* **553**, 111–114 (2018).
41. Zheng, Y. et al. Structure of CC chemokine receptor 2 with orthosteric and allosteric antagonists. *Nature* **540**, 458–461 (2016).
42. Oswald, C. et al. Intracellular allosteric antagonism of the CCR9 receptor. *Nature* **540**, 462–465 (2016).
43. Liu, X. et al. Mechanism of intracellular allosteric β_2AR antagonist revealed by X-ray crystal structure. *Nature* **548**, 480–484 (2017).
44. Nolte, W. M. et al. A potentiator of orthosteric ligand activity at GLP-1R acts via covalent modification. *Nat. Chem. Biol.* **10**, 629–631 (2014).
45. van der Westhuizen, E. T., Valant, C., Sexton, P. M. & Christopoulos, A. Endogenous allosteric modulators of G protein–coupled receptors. *J. Pharmacol. Exp. Ther.* **353**, 246–260 (2015).
46. Hay, D. L. & Pioszak, A. A. Receptor activity-modifying proteins (RAMPs): new insights and roles. *Annu. Rev. Pharmacol. Toxicol.* **56**, 469–487 (2016).
47. Pert, C. B. & Snyder, S. H. Properties of opiate-receptor binding in rat brain. *Proc. Natl Acad. Sci. USA* **70**, 2243–2247 (1973).
48. Katritch, V. et al. Allosteric sodium in class A GPCR signaling. *Trends Biochem. Sci.* **39**, 233–244 (2014).
49. Liu, W. et al. Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **337**, 232–236 (2012).
This 1.80 Å resolution structure of the A_{2A}R was the first to directly identify the location of a conserved allosteric sodium site on a GPCR
50. Miller-Gallacher, J. L. et al. The 2.1 Å resolution structure of cyanopindolol-bound β_1-adrenoceptor identifies an intramembrane Na^+ ion that stabilises the ligand-free receptor. *PLoS One* **9**, e92727 (2014).
51. Hanson, M. A. et al. A specific cholesterol binding site is established by the 2.8 Å structure of the human β_2-adrenergic receptor. *Structure* **16**, 897–905 (2008).
52. Gimpl, G. Interaction of G protein coupled receptors and cholesterol. *Chem. Phys. Lipids* **199**, 61–73 (2016).
53. Manna, M. et al. Mechanism of allosteric regulation of β_2-adrenergic receptor by cholesterol. *eLife* **5**, e18432 (2016).
54. Cherezov, V. et al. High-resolution crystal structure of an engineered human β_2-adrenergic G protein-coupled receptor. *Science* **318**, 1258–1265 (2007).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

55. Hua, T. et al. Crystal structures of agonist-bound human cannabinoid receptor CB1. *Nature* **547**, 468–471 (2017).
56. Zhang, J. et al. Agonist-bound structure of the human P2Y12 receptor. *Nature* **509**, 119–122 (2014).
57. Huang, P. et al. Cellular cholesterol directly activates Smoothened in Hedgehog signaling. *Cell* **166**, 1176–1187 (2016).
58. Byrne, E. F. X. et al. Structural basis of Smoothened regulation by its extracellular domains. *Nature* **535**, 517–522 (2016).
59. Luchetti, G. et al. Cholesterol activates the G-protein coupled receptor Smoothened to promote Hedgehog signaling. *eLife* **5**, e20304 (2016).
60. Manglik, A., Kobilka, B. K. & Steyaert, J. Nanobodies to study G Protein-coupled receptor structure and function. *Annu. Rev. Pharmacol. Toxicol.* **57**, 19–37 (2017).
61. Rasmussen, S. G. et al. Crystal structure of the β2 adrenergic receptor–G_s protein complex. *Nature* **477**, 549–555 (2011).
62. Shukla, A. K. et al. Structure of active β-arrestin-1 bound to a G-protein-coupled receptor phosphopeptide. *Nature* **497**, 137–141 (2013).
63. Zhang, Y. et al. Cryo-EM structure of the activated GLP-1 receptor in complex with a G protein. *Nature* **546**, 248–253 (2017).
64. Liang, Y. L. et al. Phase-plate cryo-EM structure of a class B GPCR-G-protein complex. *Nature* **546**, 118–123 (2017).
65. Gurbel, P. A. et al. Cell-penetrating pepducin therapy targeting PAR1 in subjects with coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.* **36**, 189–197 (2016).
66. Kahsai, A. W. et al. Conformationally selective RNA aptamers allosterically modulate the β_2-adrenoceptor. *Nat. Chem. Biol.* **12**, 709–716 (2016).
67. Hutchings, C. J., Koglin, M., Olson, W. C. & Marshall, F. H. Opportunities for therapeutic antibodies directed at G-protein-coupled receptors. *Nat. Rev. Drug Discov.* **16**, 661 (2017).
68. Maguire, M. E., Van Arsdale, P. M. & Gilman, A. G. An agonist-specific effect of guanine nucleotides on binding to the beta adrenergic receptor. *Mol. Pharmacol.* **12**, 335–339 (1976).
69. De Lean, A., Stadel, J. M. & Lefkowitz, R. J. A ternary complex model explains the agonist-specific binding properties of the adenylate cyclase-coupled beta-adrenergic receptor. *J. Biol. Chem.* **255**, 7108–7117 (1980).
70. Gurevich, V. V., Pals-Rylaarsdam, R., Benovic, J. L., Hosey, M. M. & Onorato, J. J. Agonist-receptor-arrestin, an alternative ternary complex with high agonist affinity. *J. Biol. Chem.* **272**, 28849–28852 (1997).
71. Smith, J. S., Lefkowitz, R. J. & Rajagopal, S. Biased signalling: from simple switches to allosteric microprocessors. *Nat. Rev. Drug Discov.* **17**, 243–260 (2018).
72. Edelstein, S. J. & Changeux, J. P. Biased allostery. *Biophys. J.* **111**, 902–908 (2016).
73. Flock, T. et al. Selectivity determinants of GPCR–G-protein binding. *Nature* **545**, 317–322 (2017).
74. Gurevich, E. V., Tesmer, J. J., Mushegian, A. & Gurevich, V. V. G protein-coupled receptor kinases: more than just kinases and not only for GPCRs. *Pharmacol. Ther.* **133**, 40–69 (2012).
75. He, Y. et al. Molecular assembly of rhodopsin with G protein-coupled receptor kinases. *Cell Res.* **27**, 728–747 (2017).
76. Komolov, K. E. et al. Structural and functional analysis of a β_2-adrenergic receptor complex with GRK5. *Cell* **169**, 407–421 (2017).
77. Gurevich, V. V. & Gurevich, E. V. The molecular acrobatics of arrestin activation. *Trends Pharmacol. Sci.* **25**, 105–111 (2004).
78. Shukla, A. K. et al. Visualization of arrestin recruitment by a G-protein-coupled receptor. *Nature* **512**, 218–222 (2014).
79. Kim, Y. J. et al. Crystal structure of pre-activated arrestin p44. *Nature* **497**, 142–146 (2013).
80. Kang, Y. et al. Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* **523**, 561–567 (2015).
81. Zhou, X. E. et al. Identification of phosphorylation codes for arrestin recruitment by G protein-coupled receptors. *Cell* **170**, 457–469 (2017).
82. Kim, J. et al. Functional antagonism of different G protein-coupled receptor kinases for β-arrestin-mediated angiotensin II receptor signaling. *Proc. Natl Acad. Sci. USA* **102**, 1442–1447 (2005).
83. Thomsen, A. R. et al. GPCR–G protein–β-arrestin super-complex mediates sustained G protein signaling. *Cell* **166**, 907–919 (2016).
84. Wacker, D. et al. Structural features for functional selectivity at serotonin receptors. *Science* **340**, 615 (2013).
85. Wang, C. et al. Structural basis for molecular recognition at serotonin receptors. *Science* **340**, 610 (2013).
These studies[84,85] describe crystal structures of 5-HT_{1B} and 5-HT_{2B} GPCRs, which provide the first direct structural insights into proximal molecular triggers of biased agonism
86. Wacker, D. et al. Crystal structure of an LSD-bound human serotonin receptor. *Cell* **168**, 377–389 (2017).
87. Warne, T., Edwards, P. C., Leslie, A. G. & Tate, C. G. Crystal structures of a stabilized β_1-adrenoceptor bound to the biased agonists bucindolol and carvedilol. *Structure* **20**, 841–849 (2012).
88. Nuber, S. et al. β-Arrestin biosensors reveal a rapid, receptor-dependent activation/deactivation cycle. *Nature* **531**, 661–664 (2016).
89. Lee, M.-H. et al. The conformational signature of β-arrestin2 predicts its trafficking and signalling functions. *Nature* **531**, 665–668 (2016).
These papers[88,89] describe BRET biosensor studies that show evidence for biased agonism arising at the level of the transducer
90. Lane, J. R., May, L. T., Parton, R. G., Sexton, P. M. & Christopoulos, A. A kinetic view of GPCR allostery and biased agonism. *Nat. Chem. Biol.* **13**, 929–937 (2017).
91. Klein Herenbrink, C. et al. The role of kinetic context in apparent biased agonism at GPCRs. *Nat. Commun.* **7**, 10842 (2016).
92. Furness, S. G. et al. Ligand-dependent modulation of G protein conformation alters drug efficacy. *Cell* **167**, 739–749 (2016).
This study provides evidence supporting a role for ligand–GPCR interactions in promoting conformational selection at the level of the G protein as an unappreciated molecular mechanism underlying drug efficacy
93. Gregorio, G. G. et al. Single-molecule analysis of ligand efficacy in β_2AR–G-protein activation. *Nature* **547**, 68–73 (2017).
94. Liang, Y. L. et al. Phase-plate cryo-EM structure of a biased agonist-bound human GLP-1 receptor–G_s complex. *Nature* **555**, 121–125 (2018).
95. Cai, Y. et al. Purification of family B G protein-coupled receptors using nanodiscs: Application to human glucagon-like peptide-1 receptor. *PLoS One* **12**, e0179568 (2017).
96. Whorton, M. R. et al. A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proc. Natl Acad. Sci. USA* **104**, 7682–7687 (2007).
This key pharmacology study shows that a monomeric class A GPCR can activate and signal through G proteins
97. Farran, B. An update on the physiological and therapeutic relevance of GPCR oligomers. *Pharmacol. Res.* **117**, 303–327 (2017).
98. Harikumar, K. G., Ball, A. M., Sexton, P. M. & Miller, L. J. Importance of lipid-exposed residues in transmembrane segment four for family B calcitonin receptor homo-dimerization. *Regul. Pept.* **164**, 113–119 (2010).
99. Pin, J. P. & Bettler, B. Organization and functions of mGlu and $GABA_B$ receptor complexes. *Nature* **540**, 60–68 (2016).
100. Zhang, C. et al. Structural basis for regulation of human calcium-sensing receptor by magnesium ions and an unexpected tryptophan derivative co-agonist. *Sci. Adv.* **2**, e1600241 (2016).
101. Geng, Y. et al. Structural mechanism of ligand activation in human calcium-sensing receptor. *eLife* **5**, e13662 (2016).
102. Nuemket, N. et al. Structural basis for perception of diverse chemical substances by T1r taste receptors. *Nat. Commun.* **8**, 15530 (2017).
103. Freed, D. M. et al. EGFR ligands differentially stabilize receptor dimers to specify signaling kinetics. *Cell* **171**, 683–695 (2017).
104. Congreve, M., Oswald, C. & Marshall, F. H. Applying structure-based drug design approaches to allosteric modulators of GPCRs. *Trends Pharmacol. Sci.* **38**, 837–847 (2017).
105. Roth, B. L., Irwin, J. J. & Shoichet, B. K. Discovery of new GPCR ligands to illuminate new biology. *Nat. Chem. Biol.* **13**, 1143–1151 (2017).
106. Korczynska, M. et al. Structure-based discovery of selective positive allosteric modulators of antagonists for the M_2 muscarinic acetylcholine receptor. *Proc. Natl. Acad. Sci. USA* **115**, E2419–E2428 (2018).

Acknowledgements We are grateful to J. Tesmer for assistance with preparation of Supplementary Table 1. This work was funded by the National Health and Medical Research Council of Australia (NHMRC) (Program Grant number APP1055134). D.M.T. and A.G. are Australian Research Council Discovery Early Career Research Award Fellows, P.M.S. is a NHMRC Principal Research Fellow and A.C. is a NHMRC Senior Principal Research Fellow.

Reviewer information *Nature* thanks A. Jazayeri, R. Lefkowitz and A. Manglik for their contribution to the peer review of this work.

Author contributions All authors contributed to the writing of the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to D.M.T. or A.C.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ARTICLE

https://doi.org/10.1038/s41586-018-0242-8

Nuclear F-actin and myosins drive relocalization of heterochromatic breaks

Christopher P. Caridi[1,3], Carla D'Agostino[1,3], Taehyun Ryu[1], Grzegorz Zapotoczny[1], Laetitia Delabaere[1], Xiao Li[1], Varandt Y. Khodaverdian[1], Nuno Amaral[1,2], Emily Lin[1], Alesandra R. Rau[1] & Irene Chiolo[1]*

Heterochromatin mainly comprises repeated DNA sequences that are prone to ectopic recombination. In *Drosophila* cells, 'safe' repair of heterochromatic double-strand breaks by homologous recombination relies on the relocalization of repair sites to the nuclear periphery before strand invasion. The mechanisms responsible for this movement were unknown. Here we show that relocalization occurs by directed motion along nuclear actin filaments assembled at repair sites by the Arp2/3 complex. Relocalization requires nuclear myosins associated with the heterochromatin repair complex Smc5/6 and the myosin activator Unc45, which is recruited to repair sites by Smc5/6. ARP2/3, actin nucleation and myosins also relocalize heterochromatic double-strand breaks in mouse cells. Defects in this pathway result in impaired heterochromatin repair and chromosome rearrangements. These findings identify de novo nuclear actin filaments and myosins as effectors of chromatin dynamics for heterochromatin repair and stability in multicellular eukaryotes.

Specialized mechanisms promote repair by homologous recombination in pericentromeric heterochromatin (hereafter, heterochromatin) while preventing aberrant recombination[1,2]. Heterochromatin constitutes about 30% of fly and human genomes[3], mostly comprising repeated DNA sequences (for example, transposons and satellite repeats[4]) and 'silent' epigenetic marks[3] (such as dimethylation or trimethylation of histone H3 at lysine 9 (H3K9me2/3) and heterochromatin protein 1 (HP1)), but is absent in budding yeast. In heterochromatin, thousands to millions of identical sequences, even from different chromosomes, can engage in ectopic recombination, presenting a serious threat to genome stability in multicellular eukaryotes[1,2,5–8].

In *Drosophila melanogaster*, heterochromatin forms a distinct nuclear domain[5,9], and aberrant recombination is prevented by relocalization of double-strand breaks (DSBs) to the nuclear periphery before Rad51 recruitment[5–8,10]. Loss of components required for relocalization (such as dPIAS SUMO E3-ligase, or the Smc5/6 SUMO E3-ligase subunits (called Qjt and Cerv in *Drosophila* and are both Nse2 homologues)) or for anchoring to the periphery (such as the nuclear pore protein Nup107 or the inner nuclear membrane proteins Koi and Spag4) results in defects in heterochromatin repair and widespread chromosome rearrangements[5,7,8]. Relocalization is likely to prevent aberrant recombination by separating damaged DNAs from similar repeats on non-homologous chromosomes, while promoting 'safe' exchanges with the sister chromosome or homologue[1,2,5–8]. A similar relocalization to outside heterochromatic 'chromocentres' occurs during homologous recombination in mouse cells in the G2 phase of the cell cycle[6,11,12]. The mechanisms that drive this notable movement remain to be identified.

Actin nucleators drive DSB relocalization

Nuclear actin filaments (F-actin) form in response to DSBs in mammalian cells and have poorly understood functions in repair[13–15]. We tested the role of actin polymerization in relocalization of heterochromatic DSBs. In *Drosophila* cells, repair sites start to leave the heterochromatin domain 10 min after the induction of DSBs with ionizing radiation (IR), resulting in a decrease in repair sites (γH2Av foci) in DAPI-bright heterochromatin and an increase at the nuclear periphery 60 min after IR[5,7]. Inhibition of actin polymerization with latrunculin B (LatB) increases the number of γH2Av foci in DAPI-bright heterochromatin 60 min after IR, without affecting the total number of foci (Extended Data Fig. 1a). Similarly, activation of the Arp2/3 actin nucleator by RNA interference (RNAi) or CK666 treatment results in more foci remaining in DAPI-bright heterochromatin and fewer reaching the nuclear periphery, consistent with relocalization defects (Fig. 1a, Extended Data Fig. 1b–e). Removal of LatB or CK666 reverses the effects (Extended Data Fig. 1f, g), ruling out permanent damage to repair pathways. RNAi of the actin nucleators Spire and Dia does not affect relocalization, revealing a specific role of Arp2/3 (Extended Data Fig. 1h). The kinetics of relocalization are comparable in mouse cells, and are similarly affected by ARP3 RNAi or treatment with LatB or CK666 (Fig. 1b, Extended Data Fig. 1i–k), suggesting that relocalization pathways are conserved.

Arp2/3 is activated by the Wiskott–Aldrich syndrome protein family (Wash, Scar, Whamy and Wasp in *Drosophila*). Depletion of Wash or Scar, but not of Whamy or Wasp, causes relocalization defects (Fig. 1c, Extended Data Fig. 1l). Depletion of Arp2/3 alone, Scar and Wash together, or Arp2/3, Scar and Wash together, results in similar relocalization defects, whereas depletion of Scar or Wash alone results in an intermediate phenotype (Fig. 1c), suggesting that Scar and Wash independently activate Arp2/3 for relocalization. Arp2/3 is not required for early repair steps (formation of foci containing Mdc1 (also known as Mu2 in *Drosophila*), Atrip (also known as mus304), Smc6 or Nse2, or suppression of Rad51 foci inside the heterochromatin domain[5,7,8]; Extended Data Fig. 2a–c), suggesting that actin nucleation mediates relocalization after resection and recruitment of the Smc5/6 complex.

Epistatic analyses place Smc5/6 and Arp2/3 in the same pathway for relocalization (Fig. 1d, Extended Data Fig. 2e), and Arp2/3 co-immunoprecipitates with the Smc5/6 complex in response to IR (Fig. 1e, Extended Data Fig. 2f), suggesting that Arp2/3 has a direct role in heterochromatin repair. Accordingly, Arp2/3 is enriched at repair foci in DAPI-bright heterochromatin 10 min after IR (before relocalization[5,7]), and most Arp2/3-containing foci colocalize with the

[1]Department of Molecular and Computational Biology, University of Southern California, Los Angeles, CA, USA. [2]Present address: Karolinska Institute, Stockholm, Sweden. [3]These authors contributed equally: Christopher P. Caridi, Carla D'Agostino. *e-mail: chiolo@usc.edu

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

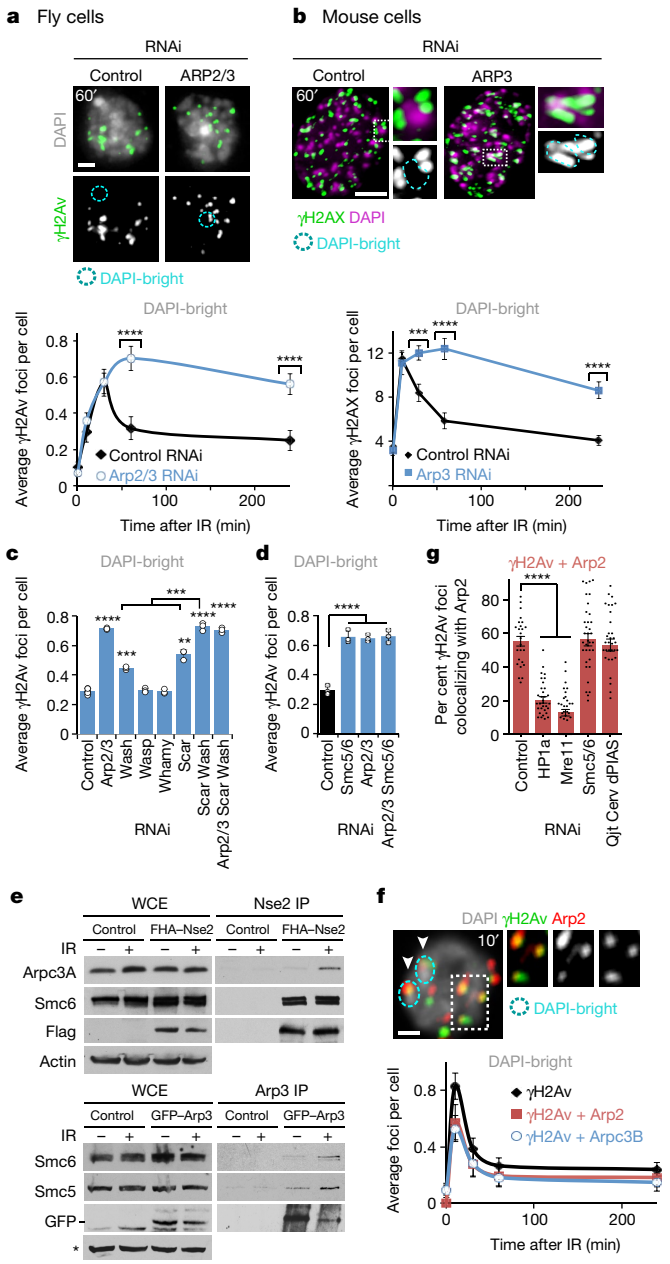


Fig. 1 | Actin nucleators mediate relocalization of heterochromatic DSBs. a, Immunofluorescence and quantification of γH2Av foci in Kc fly cells fixed at indicated time points after IR show γH2Av foci in DAPI-bright heterochromatin following indicated RNAi. $^{****}P \leq 0.0001$, $n \geq 100$ cells per RNAi per time point. **b**, As in **a**, except NIH3T3 G2 (PH3-positive) cells were used and γH2AX foci associated with DAPI-bright chromocentres were quantified. $^{****}P < 0.0001$, $^{***}P = 0.0004$, $n \geq 20$ cells per RNAi per time point. **c**, As in **a**, with immunofluorescence performed 60 min after IR. $^{****}P < 0.0001$, $^{***}P \leq 0.0009$ versus control or as indicated, $^{**}P = 0.0019$, $n > 250$ cells per RNAi. **d**, As in **c**; $^{****}P < 0.0001$, $n > 250$ cells per RNAi. **e**, Immunoprecipitation (IP) of FHA–Nse2 or GFP–Arp3 in Kc cells before (−) or after (+) IR and western blot analysis with indicated antibodies. Actin and background band (*) are loading controls. WCE, whole-cell extract. **f**, Immunofluorescence and quantification of γH2Av foci colocalizing with FHA(Flag–HA)–Arp2 or Arpc3B in DAPI-bright heterochromatin at indicated time points after IR. Arrowheads show colocalizations in DAPI-bright heterochromatin. $n > 30$ Kc cells per time point per experiment. **g**, As in **f**, showing colocalization of γH2Av and Arp2 10 min after IR, following indicated RNAi. $^{****}P < 0.0001$, $n > 22$ cells per RNAi. Scale bars, 1 μm in **a**, **f** and 5 μm in **b**. Data shown as mean ± s.e.m. in **a**, **b**, **f**, **g** and mean ± s.d. of at least three independent experiments in **c**, **d**. Headings above each graph indicate the compartments to which quantifications refer. *P* values calculated with two-tailed Mann–Whitney test.

heterochromatin mark H3K9me3 (Fig. 1f, Extended Data Fig. 2d, g, h). The Smc5/6 complex, or Smc5/6-dependent SUMOylation, might promote activation of Arp2/3 or its recruitment to DSBs. However, RNAi of Smc5/6 or SUMO E3-ligases does not affect recruitment of Arp2/3 to foci (Fig. 1g), suggesting a role for Smc5/6 in the regulation of Arp2/3. Recruitment of Arp2/3 to repair foci depends on the DSB signalling protein Mre11 and the heterochromatin component HP1a (Fig. 1g, Extended Data Fig. 2i), suggesting that the concerted activity of these factors is responsible for targeting Arp2/3 specifically to heterochromatic DSBs. We conclude that Arp2/3 is recruited to heterochromatic breaks by Mre11 and HP1a, and mediates relocalization of DSBs to the nuclear periphery in concert with Smc5/6.

Nuclear F-actin is required for relocalization

The presence of Arp2/3 at heterochromatic DSBs suggests that actin polymerizes at these sites to promote relocalization. We targeted to the nucleus a GFP-tagged F-actin chromobody (F-actCB–GFP–NLS) to visualize nuclear actin filaments in vivo without altering actin levels (Extended Data Fig. 3a). Before IR, this marker displays a diffuse nuclear signal, but IR induces the formation of striking filaments surrounding the heterochromatin domain and reaching the nuclear periphery (Fig. 2a, Extended Data Fig. 3b, Supplementary Videos 1, 2). Phalloidin staining of cells with or without the chromobody indicates that these are canonical actin filaments (Extended Data Fig. 3c, d). Consistent with a role for Arp2/3 in filament assembly, filaments are enriched for Arp2/3 (Extended Data Fig. 3e), disrupted by Arp2/3 RNAi or CK666 treatment (Fig. 2b), and characterized by branching at a $70 \pm 7°$ angle[16] (Extended Data Fig. 3f).

Filaments form concurrently with DSB departure from the heterochromatin domain (10–30 min after IR[5,7]), persist for about 25 min on average, are highly dynamic (suggesting actin remodelling[17]), and progressively disappear during relocalization of DSBs to the nuclear periphery (30–60 min after IR[7]) (Fig. 2a–c, Extended Data Figs. 3b, c, g, 4a, Supplementary Videos 2, 3). Most filaments originate from repair foci and the periphery of the heterochromatin domain (Extended Data Fig. 4b). Nearly all filaments (98.6%) remain outside the domain during elongation, suggesting that they tend to polymerize away from the domain. Filament length spans the distance between the heterochromatin domain periphery and the nuclear periphery (Extended Data Fig. 4c). Repair foci decorate actin filaments in fixed samples and travel along the filaments in time-lapse experiments (Fig. 2c, Extended Data Figs. 3d, 4a, Supplementary Video 3). Together, these data suggest that actin filaments have a direct role in relocalization of heterochromatic repair sites to the nuclear periphery.

Triton extraction partially destabilizes actin filaments, resulting in residual chromobody signals that colocalize with phalloidin and repair sites (Extended Data Fig. 3d). Similar colocalization occurs in cells expressing Flag–NLS-tagged wild-type actin (actin(WT)), but not in cells expressing a non-polymerizable R62D mutant (actin(R62D); Fig. 2d, Extended Data Fig. 4e), supporting the presence of F-actin at repair sites. Colocalization between F-actin and DSBs peaks 10–60 min after IR, followed by a marked decline (Fig. 2d). This reduction is likely to result from actin depolymerization, given that expression of a filament-stabilizing S14C mutant induces persistent colocalization (Fig. 2d, Extended Data Fig. 4e). Furthermore, consistent with a role of Arp2/3 in filament formation, colocalization between DSBs and F-actin is affected by RNAi of Arp2/3 or Scar and Wash (Extended Data Fig. 4f). Thus, Arp2/3, Scar and Wash are required to form F-actin at DSBs, with a peak of polymerization during relocalization followed by depolymerization.

We directly tested the role of nuclear F-actin in the relocalization of heterochromatic DSBs by investigating the effect of blocking actin polymerization in the nucleus without altering cytoplasmic F-actin. We expressed actin(WT) or actin(R62D) in the nucleus while blocking nuclear import of endogenous actin by RNAi depletion of the actin-specific importin Ipo9. Wild-type actin expression results in normal

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

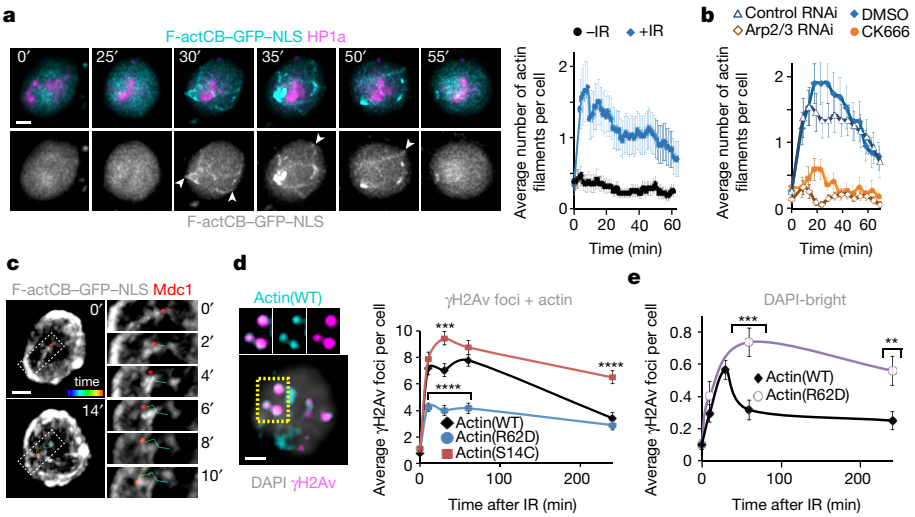


Fig. 2 | Nuclear F-actin is required for relocalization. a, Frames from Supplementary Video 2 and quantification of Kc cells expressing F-actCB–GFP–NLS and mCherry–HP1a (heterochromatin mark[5]) show dynamic filaments in cells treated (+) or not treated (−) with IR. Arrowheads: filaments reaching the nuclear periphery. Time points are after IR (frames and +IR quantification) or from the start of the video (−IR). $P < 0.0001$, IR-treated versus untreated; $n \geq 22$ cells/experiment. **b**, As in **a** but in cells treated with indicated RNAi or CK666. $P < 0.0001$, control versus Arp2/3 RNAi, DMSO versus CK666, $n \geq 19$ cells per treatment or RNAi. **c**, Frames from Supplementary Video 3 show a Mdc1 focus moving along actin filaments. Time points relative to start of focus movement. **d**, Immunofluorescence and quantification of γH2Av foci colocalizing with Flag–NLS–actin(WT), Flag–NLS–actin(R62D), or Flag–NLS–actin(S14C) at indicated time points. $^{****}P < 0.0001$, $^{***}P = 0.0006$, $n \geq 40$ cells per time point per experiment. **e**, Quantification of γH2Av foci in DAPI-bright heterochromatin at indicated time points, in cells expressing indicated actin forms after Ipo9 RNAi. $^{***}P = 0.0005$, $^{**}P = 0.0019$, $n \geq 47$ cells per time point per experiment. Scale bars, 1 μm. Data shown as mean ± s.e.m. *P* values calculated with extra sum-of-squares *F*-test, nonlinear regression for curve fitting in **a**, **b** and two-tailed Mann–Whitney test in **d**, **e**.

kinetics of focus relocalization, whereas expression of actin(R62D) leads to marked relocalization defects (Fig. 2e, Extended Data Fig. 4g–i). We conclude that nuclear actin polymerizes at repair sites, forms long filaments that reach the nuclear periphery and is required for relocalization of heterochromatic DSBs along the filaments.

Relocalization relies on nuclear myosins

Myosins I and V have been proposed to mediate nuclear dynamics along actin filaments in human cells, including for gene repositioning during transcription[18,19], chromosome territory reorganization[20] and viral infection[21]. In *Drosophila* cells, simultaneous RNAi depletion of myosin I (Myo1A and Myo1B) and myosin V (MyoV) results in defective relocalization of heterochromatic DSBs (Fig. 3a, Extended Data Fig. 5a–c). Relocalization is also impaired after nuclear expression of dominant-negative mutants of MyoV (NLS-tagged $MyoV^{headless}$ or $MyoV^{tail}$; Fig. 3a, Extended Data Fig. 5b), consistent with a nuclear function of myosins in relocalization.

Similar relocalization defects occur after RNAi depletion of Unc45 (Fig. 3a, Extended Data Fig. 5a–c), which is required for myosin stability and function. Epistasis analyses reveal that each myosin contributes independently to relocalization, and myosins and Unc45 work in the same pathway as Smc5/6 for relocalization (Fig. 3b, Extended Data Fig. 5d, e). Similar to Arp2/3 RNAi, myosin or Unc45 RNAi does not affect early repair steps (formation of foci containing Mdc1, Atrip, Smc6 or Nse2, or suppression of Rad51 focus formation inside the domain) (Extended Data Figs. 5f, g, 6a), suggesting that myosins act after resection and Smc5/6 recruitment. Myosins, and not other early repair components, co-immunoprecipitate with Smc5/6 in response to IR (Fig. 3c, Extended Data Fig. 6b, c), consistent with a direct interaction between Smc5/6 and myosins for relocalization.

Myosins and Unc45 are enriched at heterochromatic DSBs before relocalization (Fig. 3d, Extended Data Fig. 6e, f), consistent with a specific role in heterochromatin repair. Similarly to Arp2/3, recruitment of myosin to DSBs relies on Mre11 and HP1a, suggesting that the concerted activity of these factors is required to target myosin to heterochromatic DSBs (Fig. 3e). Additionally, recruitment of Unc45, but not of myosins, to DSBs relies on Smc5/6 and SUMO E3-ligases (Fig. 3e). This suggests that recruitment of Unc45 to a pre-assembled relocalization machinery (Arp2/3, F-actin and myosins) is the trigger for myosin activation and repair site relocalization downstream of Smc5/6.

The $MyoV^{headless}$ or $MyoV^{tail}$ mutations affect the ability of MyoV to walk along actin filaments, suggesting that this function is required for relocalization. Similarly, treatment of cells with the myosin inhibitor 2,3-butanedione monoxide (BDM) or the MyoV inhibitor MyoVin, which interfere with myosin movement along actin filaments, results in reversible relocalization defects (Extended Data Fig. 6g, h). BDM and MyoVin also affect relocalization in mouse cells (Extended Data Fig. 6i). We conclude that Smc5/6 interacts with nuclear myosins and recruits Unc45 to heterochromatic DSBs, and that this enables relocalization of repair sites through the ability of myosins to walk along actin filaments.

Directed motions of heterochromatic DSBs

Nuclear actin might generate filamentous structures for myosin-mediated trafficking of heterochromatic repair sites, resulting in directed motion[22]. Alternatively, F-actin and myosin could contribute to relocalization indirectly, for example by local chromatin changes enabling DNA 'looping' from the domain[23] followed by Brownian motion and anchoring at the nuclear periphery. To distinguish Brownian from directed motion, we investigated the biophysical properties of the motion by tracking Mdc1 foci and calculating the mean-square displacement (MSD)[24] of focus trajectory in *Drosophila* cells.

MSD curves for heterochromatic foci have a higher plateau than those for euchromatic foci (Fig. 4a, Extended Data Fig. 7a), showing that heterochromatic foci explore a larger space[24]. This is particularly notable considering that undamaged sites that are embedded in heterochromatin, such as centromeres, display little mobility (Fig. 4a). Heterochromatic foci that reach the nuclear periphery during the 1-h time-lapse (about 20% of all heterochromatic foci) are the most dynamic, while foci that stay inside the heterochromatin domain, move to the domain periphery or persist at the domain periphery show limited mobility (Fig. 4b, Extended Data Fig. 7b). Consistent with a specialized role for the nuclear periphery in heterochromatin repair[7], euchromatic DSBs rarely move to the nuclear periphery (Extended Data Fig. 7c).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

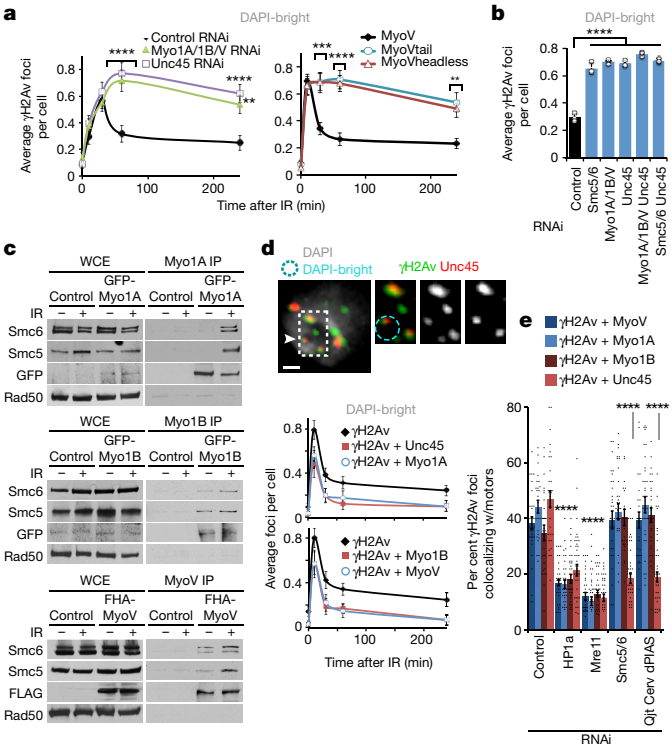


Fig. 3 | Relocalization relies on nuclear myosins. a, Quantification of γH2Av foci in DAPI-bright heterochromatin in Kc cells fixed at indicated time points after IR, and following indicated RNAi or expression of MyoV forms. ****$P < 0.0001$, ***$P \leq 0.0004$, **$P \leq 0.0033$, $n \geq 100$ cells per RNAi or cell line, per time point. **b**, As in **a**, 1 h after IR. ****$P < 0.0001$, $n \geq 468$ cells per RNAi. Control and Smc5/6 as in Fig. 1d. **c**, Immunoprecipitation of GFP–Myo1A, GFP–Myo1B or FHA–MyoV with Smc5/6 before (−) and after (+) IR and western blotting with indicated antibodies. **d**, Immunofluorescence and quantification of γH2Av foci colocalizing with FHA–Myo1A, GFP–Myo1B, MyoV, or Unc45, in DAPI-bright heterochromatin at indicated time points. $n \geq 30$ cells per time point per experiment. Scale bar, 1 μm. **e**, As in **d**, 10 min after IR, after indicated RNAi. ****$P < 0.0001$ versus control, $n > 24$ cells per experiment. Data shown as mean ± s.e.m. in **a**, **d**, **e** and mean ± s.d. of at least three independent experiments in **b**.

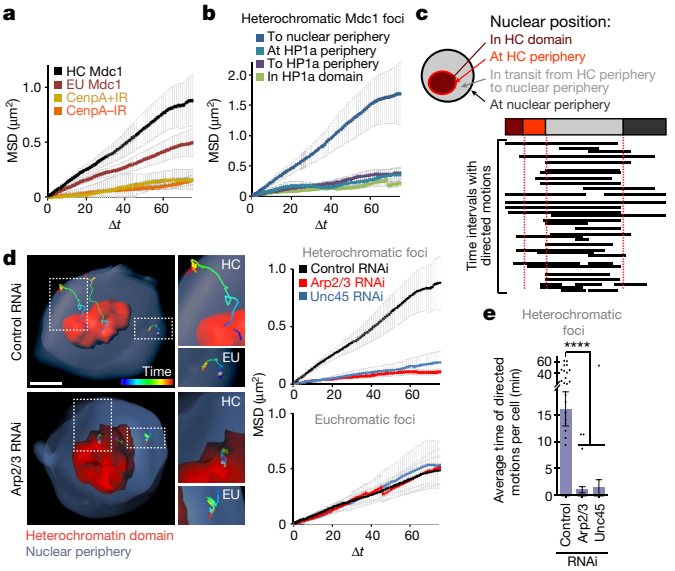


Fig. 4 | Actin nucleators and myosins drive directed motions of heterochromatic DSBs. a, MSD analysis of heterochromatic (HC) or euchromatic (EU) GFP–Mdc1 foci (formed inside or outside the mCherry–HP1a domain, respectively) after IR, or GFP–CenpA foci before (−) and after (+) IR. $P < 0.0001$ for all comparisons versus heterochromatic Mdc1 foci, $n \geq 19$ tracks per category. Δt, time intervals (intervals were 40 s each). **b**, MSD analysis of heterochromatic Mdc1 foci from **a** that remain inside the HP1a domain; move to the domain periphery; remain at the domain periphery; or move to the nuclear periphery. $P < 0.0001$ for all comparisons versus 'to nuclear periphery' category. $n = 41$ foci. **c**, LDM analysis of foci that reach nuclear periphery in **b**. Duration and length of LDMs (black lines) are adjusted relative to a 'pseudo-trajectory' defined by the average time that foci spend in each nuclear space. $n = 28$ foci. **d**, 3D reconstructions in Imaris and MSD analyses of heterochromatic foci that reach the nuclear periphery, or euchromatic foci, after indicated RNAi depletions. Heterochromatic foci: $P < 0.0001$ versus control, $n \geq 28$ heterochromatic and $n \geq 19$ euchromatic foci per RNAi. **e**, Quantification of LDM durations of all heterochromatic foci in **d**. ****$P < 0.0001$ versus control, $n \geq 30$ cells per RNAi. Scale bars, 1 μm. Data shown as mean ± s.e.m. P values calculated with extra sum-of-squares F-test, nonlinear regression for curve fitting for **a**, **b**, **d** and two-tailed Mann–Whitney test in **e**.

While MSD curves reaching a plateau typically describe Brownian motion in a confined space[24], similar plots also result from averaging MSD curves of asynchronous foci each characterized by different types of motion[22,24]. Indeed, application of a computational method[24] unmasks long-lasting directed motions (LDMs) associated with nearly all heterochromatic foci, mostly occurring between the heterochromatin domain periphery and the nuclear periphery (Fig. 4c, Extended Data Fig. 7d–f). LDMs last 24 ± 2 min (mean ± s.e.m.), consistent with the average duration of nuclear actin filaments, and are characterized by an average speed of 0.148 ± 0.01 μm min^{-1}, consistent with previously described F-actin and myosin-driven chromatin motions[18,25].

Depletion of Arp2/3 or Unc45 substantially affects the dynamics of heterochromatic foci by reducing the plateaus of MSD curves, the number of foci leaving the domain and LDM frequency (Fig. 4d, e, Extended Data Fig. 7b). Conversely, these depletions do not affect focus dynamics in euchromatin (Fig. 4d, Extended Data Fig. 7g), highlighting a specialized role for Arp2/3 and Unc45 in the directed motion of heterochromatic DSBs. Notably, Arp2/3 (but not Unc45) mediates focus clustering in euchromatin and not in heterochromatin (Extended Data Fig. 7h), revealing that Arp2/3 has distinct functions in the two domains and suggesting that focus clustering does not require extensive dynamics in *Drosophila* cells. The role of Arp2/3 and actin polymerization in focus clustering is conserved in human cells[26]. We conclude that nuclear actin filaments enable the myosin-dependent directed motion of heterochromatic repair sites to the nuclear periphery.

Importance for heterochromatin integrity

These results show that nuclear F-actin and myosins drive the relocalization of heterochromatic DSBs, suggesting that these components are important for heterochromatin repair. We directly tested this by determining the biological consequences of inactivating nuclear actin polymerization or myosins. Defective heterochromatin repair results in persistent γH2Av foci associated with the heterochromatin domain 20 h after IR in fly cells[7] and 16–24 h after IR in mouse cells[27] (that is, at time points when repair is largely completed in euchromatin[7,27]). Similar defects occur after disruption of homologous pairing (*Slmb* RNAi) or sister chromatid cohesion (*Rad21* and *Slmb* RNAi) in *Drosophila* (Extended Data Fig. 8a–c), consistent with the importance of both homologous templates for heterochromatin repair[10]. Persistent foci occur after RNAi depletion of Arp2/3, myosins and Unc45, or upon nuclear expression of a non-polymerizable form of actin in fly cells or ARP3 RNAi in mouse cells (Fig. 5a, b, Extended Data Fig. 8d, e). These depletions also impair fly cell survival after IR (Fig. 5c). We conclude that Arp2/3, nuclear F-actin and myosins are critical for the timely progression of heterochromatin repair.

Inactivation of myosins or of actin polymerization triggers the formation of IR-induced heterochromatic micronuclei in fly and mouse cells (Fig. 5d, e, Extended Data Fig. 8f, g), and genome instability in *Drosophila* larval neuroblasts (Fig. 5f, Extended Data Fig. 8h, i). Abnormal karyotypes are likely to be caused by defective

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

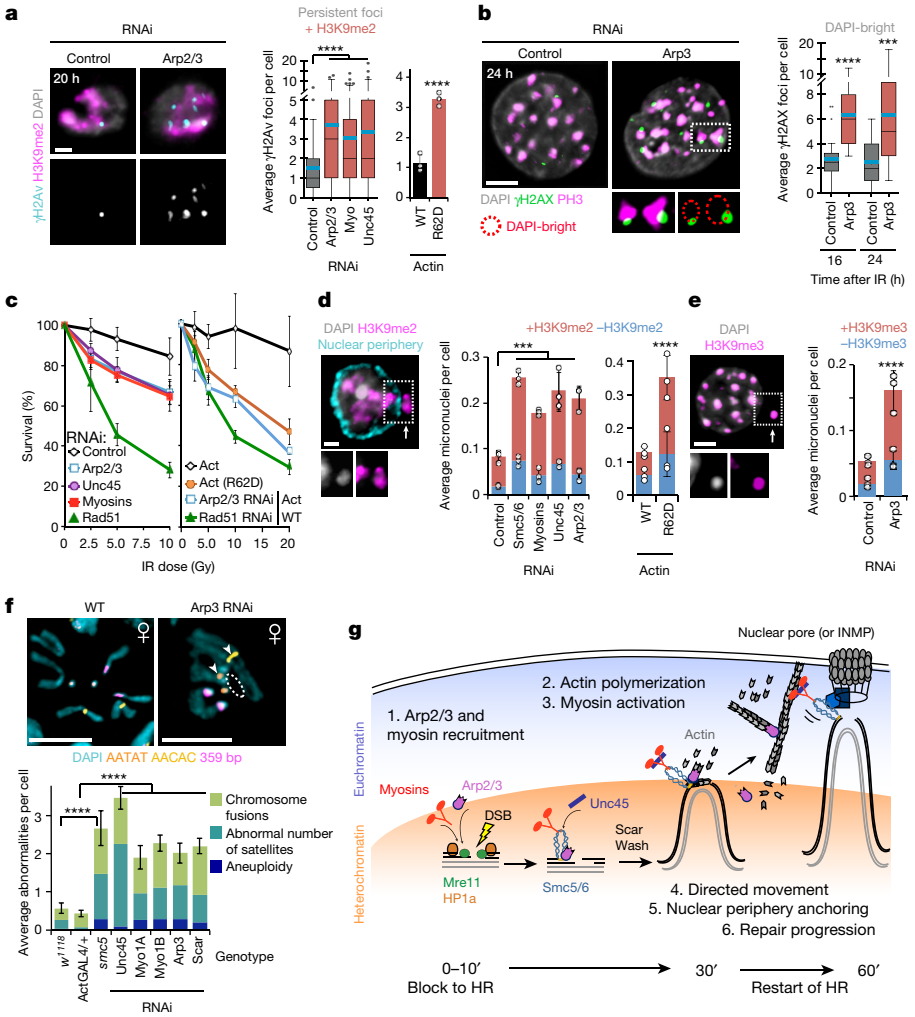


Fig. 5 | Nuclear F-actin and myosins promote heterochromatin integrity. a, Immunofluorescence and quantification of γH2Av foci associated with H3K9me2 in Kc cells fixed 20 h after IR, after indicated RNAi or expression of actin forms. $****P < 0.0001$, $n \geq 68$ cells per RNAi and $n > 220$ cells per actin-expressing line. **b**, As in **a** except NIH3T3 G2 cells stained for γH2AX and PH3. $****P < 0.0001$, $***P = 0.0008$, $n \geq 20$ cells per RNAi. **c**, IR sensitivity of Kc cells after indicated RNAi or expression of actin forms. Rad51 RNAi, positive control[7]. $P < 0.0001$ versus control RNAi or wild-type actin (Act), $n > 1{,}324$ cells per RNAi; $n \geq 248$ cells per actin-expressing line, per dose. **d**, Immunofluorescence and quantification of micronuclei (arrow) in cells stained for H3K9me2 and nuclear periphery (Nup62) after indicated RNAi or expression of actin forms and IR. $***P \leq 0.0009$, $n \geq 311$ cells per RNAi; $****P < 0.0001$, $n > 372$ cells per actin-expressing line. **e**, As in **d**, except NIH3T3 cells stained for H3K9me3. $****P < 0.0001$ versus control; $n \geq 1{,}450$ cells per RNAi. **f**, Fluorescence in situ hybridization (FISH) and quantification of larval karyotypes stained for indicated satellites show chromosome abnormalities in mutants or RNAi depleted flies versus controls (w^{1118} or ActGAL4/+). Arrowheads, chromosome fusions. Dashed circle, arm loss. $****P < 0.0001$, $n > 36$ karyotypes per genotype. Scale bars, 1 μm in **a**, **d**; 5 μm in **b**, **e**, **f**. Data shown as mean ± s.e.m. in **a** (RNAi), **b**, **f** or mean ± s.d. of at least three independent experiments in **a** (actin), **c**–**e**. *P* values calculated with two-tailed Mann–Whitney test in **a**, **b**, **d** (RNAi); unpaired *t*-test with Welch's correction in **d** (actin), **e**, **f**; extra sum-of-squares *F*-test, nonlinear regression for curve-fitting in **c**. **g**, Model of heterochromatin relocalization pathway. DSBs are resected inside the heterochromatin domain; Mre11 and HP1a promote the loading of Arp2/3 and nuclear myosins to DSBs; activation of Arp2/3 by Scar and Wash induces formation of actin filaments that reach the nuclear periphery; Unc45 recruitment by Smc5/6 activates nuclear myosins to 'walk' along filaments, thus relocalizing DSBs to nuclear pores or inner nuclear membrane proteins (INMPs)[5,7,8]; homologous recombination continues while filaments disassemble. Relocalization prevents ectopic recombination by isolating damaged sites and their homologous templates (grey lines) from undamaged heterochromatic repeats before strand invasion.

repair of spontaneous DSBs during larval development. Chromosome rearrangements include aneuploidies, chromosome fusions and changes in the number of satellites (Fig. 5f); most of which involve the heterochromatic fourth and Y chromosomes, or pericentromeric regions, as expected for defective heterochromatin repair[7] (Extended Data Fig. 8i). We conclude that Arp2/3-dependent nuclear actin polymerization and myosins are critical for heterochromatin repair and stability in *Drosophila* cells and tissues, and in mouse cells.

Discussion

Our results revealed a notable role of F-actin and myosins in the directed motion of heterochromatic DSBs to the nuclear periphery. To our knowledge, we have shown for the first time by live imaging the formation, structure and remodelling of IR-induced nuclear actin filaments and the movement of repair sites along them. Filaments form at repair sites, reach the nuclear periphery and disassemble after relocalization. We have also identified a specialized network of proteins that coordinates heterochromatin dynamics and nuclear actin filament formation in concert with Smc5/6, including Arp2/3; the Arp2/3 activators Scar and Wash; the nuclear myosins Myo1A, Myo1B and MyoV; and the myosin activator Unc45. Arp2/3 and myosins are targeted to heterochromatic DSBs by Mre11 and the heterochromatin component HP1a. Recruitment of Unc45 also requires Smc5/6 and SUMO E3-ligases, suggesting that Unc45 is a trigger for the myosin-dependent movement

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

of heterochromatic DSBs along actin filaments downstream of Smc5/6. This pathway appears to be conserved in mouse cells, and its deregulation results in heterochromatin repair defects and instability. These results support a model in which DSBs in heterochromatin induce the formation of nuclear actin filaments that connect repair sites to the nuclear periphery. Smc5/6-associated myosins travelling along these 'highways for repair' enable the directed movement of repair sites to the nuclear periphery for safe progression of homologous recombination repair (Fig. 5g).

Nuclear repositioning of repair sites has been observed in different contexts[1,2], including DSBs in rDNA[28,29], damaged telomeric and subtelomeric sequences[30–35], collapsed replication forks[36,37], persistent DSBs[36,38–40], and homology search[41,42]. Cytoplasmic microtubules contribute to some of these dynamics via the trans-nuclear membrane LINC complex[35,40]. However, the nuclear structures responsible for those movements have remained largely unknown, and Brownian motions are thought to account for most of these dynamics[23].

Directed motions might be especially needed to leave the heterochromatin domain, given the compact nature of this domain and its potential resistance to releasing repair sites. Accordingly, Arp2/3 and myosins affect directed motions in heterochromatin and not in euchromatin. Notably, Arp2/3-dependent nuclear actin filaments form loose branches, distinct from the dense cytoplasmic network that mediates cell movement. This is likely to reflect specialized Arp2/3 regulation and F-actin remodelling[17] in the nucleus for the directed motion of repair sites. These studies identify nuclear F-actin and myosins as critical components for heterochromatin repair and genome stability in multicellular eukaryotes.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0242-8.

Received: 15 June 2017; Accepted: 18 May 2018;
Published online 20 June 2018.

1. Amaral, N., Ryu, T., Li, X. & Chiolo, I. Nuclear dynamics of heterochromatin repair. *Trends Genet.* **33**, 86–100 (2017).
2. Caridi, P. C., Delabaere, L., Zapotoczny, G. & Chiolo, I. And yet, it moves: nuclear and chromatin dynamics of a heterochromatic double-strand break. *Phil. Trans. R. Soc. Lond. B* **372**, 20160291 (2017).
3. Ho, J. W. et al. Comparative analysis of metazoan chromatin organization. *Nature* **512**, 449–452 (2014).
4. Hoskins, R. A. et al. The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Res.* **25**, 445–458 (2015).
5. Chiolo, I. et al. Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. *Cell* **144**, 732–744 (2011).
6. Chiolo, I., Tang, J., Georgescu, W. & Costes, S. V. Nuclear dynamics of radiation-induced foci in euchromatin and heterochromatin. *Mutat. Res.* **750**, 56–66 (2013).
7. Ryu, T. et al. Heterochromatic breaks move to the nuclear periphery to continue recombinational repair. *Nat. Cell Biol.* **17**, 1401–1411 (2015).
8. Ryu, T., Bonner, M. R. & Chiolo, I. Cervantes and Quijote protect heterochromatin from aberrant recombination and lead the way to the nuclear periphery. *Nucleus* **7**, 485–497 (2016).
9. Li, Q. et al. The three-dimensional genome organization of *Drosophila melanogaster* through data integration. *Genome Biol.* **18**, 145 (2017).
10. Janssen, A. et al. A single double-strand break system reveals repair dynamics and mechanisms in heterochromatin and euchromatin. *Genes Dev.* **30**, 1645–1657 (2016).
11. Jakob, B. et al. DNA double-strand breaks in heterochromatin elicit fast repair protein recruitment, histone H2AX phosphorylation and relocation to euchromatin. *Nucleic Acids Res.* **39**, 6489–6499 (2011).
12. Tsouroula, K. et al. Temporal and spatial uncoupling of DNA double strand break repair pathways within mammalian heterochromatin. *Mol. Cell* **63**, 293–305 (2016).
13. Andrin, C. et al. A requirement for polymerized actin in DNA double-strand break repair. *Nucleus* **3**, 384–395 (2012).
14. Belin, B. J., Lee, T. & Mullins, R. D. DNA damage induces nuclear actin filament assembly by Formin -2 and Spire-1/2 that promotes efficient DNA repair. *eLife* **4**, e07735 (2015).
15. Sun, M. H. et al. DNA double-strand breaks induce the nuclear actin filaments formation in cumulus enclosed oocytes but not in denuded oocytes. *PLoS ONE* **12**, e0170308 (2017).
16. Mullins, R. D., Heuser, J. A. & Pollard, T. D. The interaction of Arp2/3 complex with actin: nucleation, high affinity pointed end capping, and formation of branching networks of filaments. *Proc. Natl Acad. Sci. USA* **95**, 6181–6186 (1998).
17. Goley, E. D. & Welch, M. D. The ARP2/3 complex: an actin nucleator comes of age. *Nat. Rev. Mol. Cell Biol.* **7**, 713–726 (2006).
18. Chuang, C.-H. et al. Long-range directional movement of an interphase chromosome site. *Curr. Biol.* **16**, 825–831 (2006).
19. Dundr, M. et al. Actin-dependent intranuclear repositioning of an active gene locus *in vivo*. *J. Cell Biol.* **179**, 1095–1103 (2007).
20. Mehta, I. S., Amira, M., Harvey, A. J. & Bridger, J. M. Rapid chromosome territory relocation by nuclear motor activity in response to serum removal in primary human fibroblasts. *Genome Biol.* **11**, R5 (2010).
21. Roberts, K. L. & Baines, J. D. Actin in herpesvirus infection. *Viruses* **3**, 336–346 (2011).
22. Hatakeyama, H., Nakahata, Y., Yarimizu, H. & Kanzaki, M. Live-cell single-molecule labeling and analysis of myosin motors with quantum dots. *Mol. Biol. Cell* **28**, 173–181 (2017).
23. Amitai, A., Seeber, A., Gasser, S. M. & Holcman, D. Visualization of chromatin decompaction and break site extrusion as predicted by statistical polymer modeling of single-locus trajectories. *Cell Reports* **18**, 1200–1214 (2017).
24. Caridi, C. P. et al. Quantitative methods to investigate the 4D dynamics of heterochromatic repair sites in *Drosophila* cells. *Methods Enzymol.* **601**, 359–389 (2018).
25. Li, H., Guo, F., Rubinstein, B. & Li, R. Actin-driven chromosomal motility leads to symmetry breaking in mammalian meiotic oocytes. *Nat. Cell Biol.* **10**, 1301–1308 (2008).
26. Schrank, B. R. et al. Nuclear ARP2/3 drives DNA break clustering for homology-directed repair. *Nature* https://doi.org/10.1038/s41586-018-0237-5 (2018).
27. Goodarzi, A. A. et al. ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. *Mol. Cell* **31**, 167–177 (2008).
28. Torres-Rosell, J. et al. The Smc5–Smc6 complex and SUMO modification of Rad52 regulates recombinational repair at the ribosomal gene locus. *Nat. Cell Biol.* **9**, 923–931 (2007).
29. van Sluis, M. & McStay, B. A localized nucleolar DNA damage response facilitates recruitment of the homology-directed repair machinery independent of cell cycle stage. *Genes Dev.* **29**, 1151–1163 (2015).
30. Therizols, P. et al. Telomere tethering at the nuclear periphery is essential for efficient DNA double strand break repair in subtelomeric region. *J. Cell Biol.* **172**, 189–199 (2006).
31. Khadaroo, B. et al. The DNA damage response at eroded telomeres and tethering to the nuclear pore complex. *Nat. Cell Biol.* **11**, 980–987 (2009).
32. Cho, N. W., Dilley, R. L., Lampson, M. A. & Greenberg, R. A. Interchromosomal homology searches drive directional ALT telomere movement and synapsis. *Cell* **159**, 108–121 (2014).
33. Chung, D. K. et al. Perinuclear tethers license telomeric DSBs for a broad kinesin- and NPC-dependent DNA repair process. *Nat. Commun.* **6**, 7742 (2015).
34. Churikov, D. et al. SUMO-dependent relocalization of eroded telomeres to nuclear pore complexes controls telomere recombination. *Cell Reports* **15**, 1242–1253 (2016).
35. Lottersberger, F., Karssemeijer, R. A., Dimitrova, N. & de Lange, T. 53BP1 and the LINC complex promote microtubule-dependent DSB mobility and DNA repair. *Cell* **163**, 880–893 (2015).
36. Nagai, S. et al. Functional targeting of DNA damage to a nuclear pore-associated SUMO-dependent ubiquitin ligase. *Science* **322**, 597–602 (2008).
37. Su, X. A., Dion, V., Gasser, S. M. & Freudenreich, C. H. Regulation of recombination at yeast nuclear pores controls repair and triplet repeat stability. *Genes Dev.* **29**, 1006–1017 (2015).
38. Kalocsay, M., Hiller, N. J. & Jentsch, S. Chromosome-wide Rad51 spreading and SUMO-H2A.Z-dependent chromosome fixation in response to a persistent DNA double-strand break. *Mol. Cell* **33**, 335–343 (2009).
39. Oza, P., Jaspersen, S. L., Miele, A., Dekker, J. & Peterson, C. L. Mechanisms that regulate localization of a DNA double-strand break to the nuclear periphery. *Genes Dev.* **23**, 912–927 (2009).
40. Swartz, R. K., Rodriguez, E. C. & King, M. C. A role for nuclear envelope-bridging complexes in homology-directed repair. *Mol. Biol. Cell* **25**, 2461–2471 (2014).
41. Dion, V., Kalck, V., Horigome, C., Towbin, B. D. & Gasser, S. M. Increased mobility of double-strand breaks requires Mec1, Rad9 and the homologous recombination machinery. *Nat. Cell Biol.* **14**, 502–509 (2012).
42. Miné-Hattab, J. & Rothstein, R. Increased chromosome mobility facilitates homology search during recombination. *Nat. Cell Biol.* **14**, 510–517 (2012).

Acknowledgements Work supported by USC Gold Family and Research Enhancement Fellowships to T.R., and R01GM117376, Mallinckrodt Foundation Award, and NSF Career 1751197 to I.C. We thank S. Keagy, O. Aparicio and the Chiolo Laboratory for comments; O. Aparicio, M. Michael and V. Longo for reagents and cells; B. Mellone, G. Posern and S. Celniker for plasmids; L. Cooley, A. Eprussi, H. Ohkura, M. Gatti, S. Bernstein, G. Karpen, J. Kadonaga, S. Elgin, S. Parkhurst and DSHB for antibodies; F. Cuturale, C. Artesano, S. Fraser and the Translation Imaging Center for help with imaging and analyses; H. Hopp, B. Spatola, P. Palagani and A. Holcom for experimental help; and the BDSC (P40OD018537), VDRC, and DGRC (P40OD010949) for flies and plasmids. We apologize to our colleagues whose work could not be cited owing to space limitations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Author contributions C.P.C. and C.D. performed most experiments; T.R. performed experiments for Figs. 1e, 3a, c, 5c–e and Extended Data Figs. 1f, g, 2i, 4i, 5d, 6c, h, 8c; G.Z. for Figs. 1b, 3a, 5b, e and Extended Data Figs. 1k, 2f, 6i; L.D. for Extended Data Figs. 3e, 8h and complementary studies; X.L. for Fig. 5f and Extended Data Fig 8h; V.Y.K for Fig. 3a and Extended Data Figs. 1g, 6b, h, 8a–c, f; N.A. for Fig. 2a, d; E.L. for Fig. 5c; and A.R.R. for Extended Data Fig. 8h. I.C. conceived and supervised the project, contributed to experiments and wrote the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0242-8.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0242-8.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to I.C.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Cell culture and generation of stable cell lines. Kc167 (Kc) cells were used for most experiments and were maintained as logarithmically growing cultures in Schneider's medium (Sigma) + FBS (Gemini). Kc cells were authenticated by the *Drosophila* Genomic Resource Center (DGRC) and no mycoplasma contamination was detected[43]. Stable lines were obtained by cotransfecting the plasmid of interest with pCoHygro (Invitrogen) or pCoPuro (Addgene) and selecting in the presence of 100 μg/ml hygromycin B (Invitrogen) or 2 mg/ml puromycin (Enzo Life Sciences). Transfection was performed with DOTAP (Roche) or Cellfectin (Life Technologies), according to the manufacturers' protocols. Mouse NIH3T3 cells (gift from V. Longo) were maintained at 30–80% confluency in DMEM (Corning) + 10% CBS (Denver Serum Company).

IR treatments. Cell cultures were exposed to IR using a 160 kV X-ray source (X-RAD iR-160, Precision X-Ray). For *Drosophila* cells, we mostly use a range of Gy at which the damage response increases linearly with dose (1.7–20 Gy), and corresponds to nearly sublethal doses for controls[7]. A dose of 5 Gy was used for most experiments, unless otherwise indicated. Treatments with this dose do not affect nuclear–cytoplasmic transport, which was evaluated by measuring the ratio between nuclear and cytoplasmic GFP-tagged Myo1A, Myo1B, MyoV, Arp2, Arp3 and Unc45. The estimated number of DSBs induced by 5 Gy in *Drosophila* cells is ~7.5 DSBs in G1 and ~14 DSBs in G2[7]. In kinetic analyses of fixed cells, time 0 corresponds to cells fixed without exposure to IR. In time-lapse experiments, time 0 corresponds to cells imaged 5–10 min before IR treatment unless otherwise specified. Mouse cells were exposed to 1 Gy, similar to previous studies[27,44,45].

Plasmids. pCopia–GFP–Mdc1, pCopia–mCherry–HP1a, pCopia–GFP–Smc5, pCopia–GFP–Nse2, pCopia–GFP–Atrip, and pCopia–GFP–CenpA plasmids were previously described[5,7,46]. All other GFP- and FHA (3×Flag–3×HA)-tagged proteins were generated by insertion of PCR-amplified coding regions of cDNAs from DGRC or other sources. Clone numbers from DGRC were: Arp2 (LD18955), Arp3 (LD35711), Myo1A (SD01662), Myo1B (FI19407) and Unc45 (SD10334). Detailed information is available on the DGRC website (http://dgrc.cgb.indiana.edu). Arp2/3 has previously been identified as a major actin-nucleating factor in *Drosophila*[47]. Functions of Unc45 in myosin regulation have been described[48]. Wild-type actin and the two mutant constructs actin(S14C) and actin(R62D) were PCR amplified from pEF–Flag–NLS–βActin, pEF–Flag–NLS–βActin(S14C) or pEF–Flag–NLS–βActin(R62D)[49] (gift from G. Posern). MyoV was generated by PCR amplification of clone FI23904 (ref. 2014422, gift from S. Celniker). The actin chromobody was generated by PCR amplification of actin-chromobody–TagGFP2–NLS plasmid[50]. MyoV^{headless} and MyoV^{tail} mutants were generated by PCR amplifications of truncated versions of MyoV, according to ref. [51]. All PCR products were cloned into pCopia–LAP–EGFP vectors[52], or pCopia–3×Flag–StrepII–3×HA vectors[5]. pCopia–mCherry–Mdc1 was generated by swapping the *Mdc1* gene from pCopia–GFP–Mdc1 into a pCopia–mCherry vector[5].

dsRNA synthesis and sequences. siRNAs for RNAi depletions in *Drosophila* cultured cells were prepared with the MEGAscript T7 Kit (Applied Biosystems). dsRNA derived from the *brown* (*bw*) gene was used as control in all experiments. Amplicons and RNAi conditions for *bw*, Smc5, Smc6 and Rad51 were as previously described[5]. Amplicons used for all other dsRNAs were: DRSC25363/DRSC34284 for Myo1A, DRSC27633 for Myo1B, DRSC07603/DRSC40837 for MyoV, DRSC15449 for Unc45, DRSC31428/DRSC19332 for Arp2, DRSC29666/DRSC31430 for Arp3, DRSC03426/BKN22532 for Scar, DRSC37426/DRSC06302 for Wash, DRSC26170/DRSC15032 for Wasp, DRSC03519/DRSC24127 for Dia, DRSC03619/DRSC24744/DRSC40746 for Spire, DRSC40055 for Ipo9, DRSC17056/DRSC32610 for Slmb and DRSC20839/DRSC38781 for Rad21. Sequences can be found on the DRSC website (http://flyrnai.org). dsRNA for Whamy RNAi was prepared with the oligonucleotides: 5′GAATTAATACGACTCACTATAGGGAGACGGAATACGGATTATCCGCC and 5′ GAATTAATACGACTCACTATAGGGAGAGATGCAGAAATCACAGCCCAA. Wash, Scar, Whamy and Wasp are the previously described Wiskott–Aldrich Syndrome protein family in flies[53]. When more than one amplicon is indicated, we combined equal amounts of each dsRNA for better efficiency of protein depletion. For mouse cells, RNAi experiments were conducted using SMARTpool ON-TARGETplus ARP3 siRNA (Dharmacon) and a corresponding non-targeting RNAi control.

RNAi depletion. For *Drosophila* cells, dsRNAs were transfected with DOTAP (Roche) following the manufacturer's instructions and depletion was done for 5 days before performing experiments, except for Slmb and Rad21 RNAi, for which this was done for 4 days. We did not detect significant effects of the RNAi depletions shown in this paper on cell cycle progression by FACS, or on the heterochromatin domain by live imaging of GFP–HP1a-expressing cells. Additionally, RNAi depletion of Arp2/3, myosins or Unc45 did not affect nuclear pore structure by Nup62 immunofluorescence staining. Notably, studies in flies have shown that loss of Wash affects association of HP1a with heterochromatic sequences[54]. However, depletion of Wash in Kc cells did not affect the mCherry–HP1a signal, indicating that the heterochromatin domain was still largely intact in the RNAi conditions used in Fig. 1c and Extended Data Figs. 1b, l, 4f. This supports a model in which Wash works downstream of HP1a in the relocalization of heterochromatic DSBs. We also note that all kinetics resulting from RNAi depletion must be compared to cells treated with control dsRNAs, as the γH2Av peak shifts from 10 min after IR in non-RNAi experiments (for example, Fig. 1f) to 30 min after IR in RNAi controls[5] (for example, Fig. 1a). RNAi depletion efficiencies for Smc5/6, Nse2/Cerv, Nse2/Qjt, dPIAS, HP1a and Rad51 have previously been validated[5,7,8]. Mouse cells were transfected using Lipofectamine RNAiMAX reagent (Thermo Fisher Scientific) for 2 days before the experiments.

Quantitative PCR. qPCR analysis of RNAi depletions in *Drosophila* cultured cells was performed as described[7]. Changes in transcript levels were normalized to *Ebony*, *CG12493* or *Actin5c* mRNA. Primer sequences were: CGCACGGAAATTATCAAGGT and GTCTGCTCGTCATCCTCCTC for Ipo9; ATCATTCACTCCAATTTTACT and AGTAATCTGGTTATTAGTGGAA for Whamy, GGCTGAAGAAGAGCAACCAC and TCACCTTGCGCAACTGACCAG for Spire; GCGAGTTGTTCCCCAAGATA and GGAGACAATCTTGCCGTCAT for Slmb. For qPCR analysis of RNAi depletions in flies, total RNA was isolated from 5–10 larvae by Trizol extraction and treated with DNase I to remove genomic DNA. RNA was used to generate single-stranded cDNA using oligo dT priming and Superscript Reverse Transcriptase III (Invitrogen). Primer sequences were: CTTAACGAAGACCTGAATGAGGC and CGACAGTGTCTCCTGTTCCAG for Lamin; GAATGCCCCACATTTGTTCGC and ACCGGAAATTAGGATGCAGGT for Myo1A; AGAGCGAAGAAGCCTTCATCG and GCTTGTAGGGATTCACGGAGAT for Myo1B. Transcripts were quantified with iQ SYBR Green Supermix (Bio-Rad) according to the manufacturer's instruction. Changes in transcript levels were normalized to *Lamin* mRNA. In both cultured cells and flies, each qPCR was repeated at least three times, and graphs show the average level of depletion relative to control RNAi. qPCR analysis of *Arp3* RNAi depletion in mouse cells was done using two independent primer sets for both *Arp3* and the control mRNA *Gapdh*, and the qPCR was repeated twice for each set. Averaged values were normalized for each GAPDH qPCR and averaged between each other. The primer sequences used for this analysis were: AAGAATGAAGCGGACAGGAC and AGTTGGTGATTCCTAGCGTG (set 1) AATCACCAACTTCCCCATCC and AAACCTTCTCACACAGCCTC (set 2) for Arp3; CTTTGTCAAGCTCATTTCCTGG and TCTTGCTCAGTGTCCTTGC (set 1) and GCCTTCCGTGTTCCTACC and CCTCAGTGTAGCCCAAGATG (set 2) for GAPDH.

Immunofluorescence and quantification in fixed samples. Chromosome preparation and FISH protocols used in Fig. 5f and Extended Data Fig. 8b, i were as previously described[7,9,55]. AACAC, AATAT and 359-bp probes were designed as previously described[55] and were purchased from Integrated DNA Technologies. Probe sequences are: 5′-6-FAM-$(AACAC)_7$, 5′-Cy3-TTTTCCAAATTTCGGTCATCAAATAATCAT, and 5′-Cy5-$(AATAT)_6$, respectively. Immunofluorescence without triton extraction was used for most experiments as previously described[5]. Immunofluorescence staining of myosins, Unc45, Arp2/3 complex subunits, Flag–NLS–actin (actin(WT), actin(R62D) and actin(S14C) in Fig. 2d and Extended Data Fig. 4d, e), and F-actCB–GFP–NLS in Extended Data Fig. 3d, was preceded by a triton extraction step as previously described[7]. Notably, these Triton extraction treatments partially destabilize actin filaments[56–58]. Detection of nuclear actin filaments with phalloidin in Extended Data Fig. 3c was done by fixing cells for 10 min at 37 °C in pre-warmed, freshly made, fixation buffer (4% PFA, 0.5% Triton-X100, 2 mM ATP, 0.004% NaAz, 1 mM Na3VO4, 1 mM NaF, 1 mM PMSF, protease inhibitor tablets (Roche), phalloidin 1:500; modified from published protocols[59,60]). Imaging and image processing for fixed cells and tissues has previously been described[5,7]. Quantification of repair foci relative to DAPI-bright in *Drosophila* cells was done as previously described[5,7]. Quantification of repair foci relative to H3K9me2 in *Drosophila* cells was done as previously described[7]. Quantification of repair foci relative to DAPI-bright in mouse cells was done on volume reconstructions and only cells in G2 were used; detection of G2 cells was based on PH3 staining as previously described[44]. For quantification of actin filaments in Extended Data Fig. 3c, filaments were detected with phalloidin (1:500 Alexa Fluor 488, Invitrogen, A12379), imaged with 0.2-μm intervals between the *z* stacks, and images were deconvolved 20 times with a conservative protocol. Filaments inside the nuclei were identified using DAPI staining as a reference. Figures shown are maximum intensity projections of a few *z* stacks across the heterochromatin domain(s) (Figs. 1a, b, 5a, b; Extended Data Fig. 1a), individual *z* stacks or maximum intensity projections of 2–3 *z* stacks (Figs. 1f, 2d, 3d, 5f, Extended Data Figs. 1d, 2g, 3c, d, e, 4b, d, e, 6e, 8i), maximum intensity projections of all *z* stacks across the nucleus (Fig. 5d, e, Extended Data Fig. 8b), or volume reconstructions (Figs. 2a, 4d, Extended Data Figs. 2a, c, 3b, f, 4a, 5f, g, 7a).

Western blotting. Protein extraction and western blotting to validate RNAi depletions were conducted as previously described[7].

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Immunoprecipitation. Immunoprecipitation experiments were performed as previously described[7]. For Myo1A immunoprecipitation, 3×10^8 Kc cells expressing GFP–Myo1A, or Kc cells as a control, were pelleted and snap-frozen before and 25 min after exposure to 20 Gy IR. Pellets were incubated in lysis buffer A (50 mM Hepes, 10 mM KCl, 2 mM MgCl2, 20% glycerol, 1 mM PMSF, 20 mM NaF, 20 mM glycerol-2-phosphate, 1 mM benzamidine, 0.5% NP-40, 25 mM NEM, 150 mM NaOAc, 250 mM NaCl, and protease inhibitors (Complete, Roche)), and digested with 1 μl benzonase (EMD Millipore) at 4 °C for 30 min. Pellets were centrifuged and soluble lysates collected. Pellets were resuspended in buffer A and extracted with 300 mM sodium acetate for 1 h at 4 °C. Next, the supernatant was pooled with the soluble lysate. Samples were incubated for 3 h at 4 °C with protein G-coupled sepharose beads and 5 μl goat anti-GFP antibody. Beads were washed 5 times with wash buffer A (50 mM Hepes, 10 mM KCl, 2 mM $MgCl_2$, 20% glycerol, 0.5% NP-40, 350 mM NaCl, 150 mM NaOAc, and 25 mM NEM) and heated in loading buffer for 3 min at 95 °C for SDS–PAGE.

For Myo1B, Arp3 and Nse2 immunoprecipitation, 3×10^8 Kc cells expressing GFP–Myo1B, GFP–Arp3 or FHA–Nse2 were pelleted and snap-frozen before and 25 min after exposure to 20 Gy IR. Pellets were incubated with lysis buffer C (50 mM Hepes/KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.05% Na-deoxycholate, 25 mM NEM) for 30 min at 4 °C. Lysates were pelleted and the supernatant was incubated with 1 μl of benzonase (EMD Millipore) for 30 min at 4 °C. Next, supernatants were incubated with Flag–M2 agarose beads (Sigma) for Nse2 or protein G-coupled sepharose beads and 5 μl goat anti-GFP antibody for Myo1B for 4 h at 4 °C. Samples were washed once with lysis buffer C and four times with wash buffer B (100 mM Tris pH 8.0, 250 mM NaCl, 1 mM EDTA, 0.5% NP-40, 0.1% Na- deoxycholate), before heating them in loading buffer for 3 min at 95 °C for SDS–PAGE. As a control for Myo1B immunoprecipitation, GFP–Myo1B cells were incubated with sepharose beads without antibodies. Kc cells were used as control for Nse2 and Arp3 immunoprecipitation.

For MyoV immunoprecipitation, 3×10^8 Kc cells expressing FHA–MyoV, or cells without the FHA–MyoV plasmid as a control, were used. Cells were pelleted and snap-frozen before and 25 min after exposure to 20 Gy IR. Pellets were incubated in lysis buffer B (50 mM Tris, 150 mM NaCl, 2 mM EDTA, 10% glycerol, 0.2% Nonidet P-40, protease inhibitors (Complete, Roche), 1 mM PMSF, and 25 mM NEM) for 30 min at 4 °C. Lysates were centrifuged and the supernatant was incubated with Flag–M2 agarose beads (Sigma). Samples were incubated with beads at 4 °C for 5 h. Beads were washed 5 times with lysis buffer B and heated in loading buffer for 3 min at 95 °C for SDS–PAGE.

Cytoplasmic and nuclear actin fractionation. Protocols were adapted from published methods[61]. All steps were carried out in ice or in a cold room. First, 10^7 cells were collected, washed once with ice-cold phosphate-buffered saline (PBS), and pelleted by centrifugation at 800*g* for 10 min. Cell pellets were placed at −80 °C for 1 h. Then, pellets were resuspended gently on ice in 200 μl buffer P1 (HEPES 10 mM, EGTA 0.1mM, DTT 1 mM, complete protease inhibitors (Roche)). After addition of Triton X-100 (final concentration 0.5%), samples were vortexed on the highest setting for 10 s followed by centrifugation at 4 °C, 10,000*g* for 10 min. One hundred and fifty microlitres of supernatant was collected for the cytoplasmic fraction. The remaining supernatant was removed and discarded. The cytoplasmic fraction was centrifuged at 16,000*g*, transferred to a new tube, and placed at −80 °C. The nuclear pellet was gently resuspended in 1 ml buffer P1 followed by centrifugation at 4 °C, 10,000*g* for 10 min. The supernatant of the nuclear fraction was discarded and nuclear pellet lysed in 50 μl buffer P2 (20 mM HEPES, 25% glycerol, 400 mM NaCl, 1 mM EGTA, 1 mM DTT, complete protease inhibitors (Roche)) by vortexing for 90 min at 4 °C. Nuclear extract was clarified by centrifugation at 20,000*g* for 10 min. The samples were stored at −80 °C until SDS–PAGE and western blotting were performed. Twenty-five micrograms of protein extract (nuclear or cytoplasmic) was loaded onto an SDS–PAGE gel, per lane.

Chemical treatments. Myosin and actin polymerization inhibitors were added to the cultures before IR treatments. In *Drosophila* cells, treatment conditions were: 2 mM and 10 mM BDM[62] (Sigma) for 5 min; 200 and 400 μM MyoVin-1[63] (Calbiochem) for 2 h; 100 μM and 400 μM CK666[64] (Sigma) for 60 min; 4 μM and 8 μM latrunculin B[65] (Enzo Life Sciences) for 20 min. Stocks of MyoVin-1 (100 mM), CK666 (100 mg/ml) and latrunculin B (10 mM) were prepared in DMSO and those of BDM were diluted in water. For mouse cells, treatment conditions were: 7.5 mM BDM for 20 min; 200 μM MyoVin-1 for 60 min; 400 μM CK666 for 60 min; 8 μM latrunculin B for 20 min. All stocks were prepared in DMSO. For controls, DMSO or water was added to the medium to yield final concentrations equivalent to samples treated with chemicals. For the 'release' time points shown in Extended Data Figs. 1f, g, 6h, Kc cells were washed 3 times after incubation with each chemical, and incubated at 27 °C for 1 h before IR. Cells were fixed 60 min after IR. Notably, the doses of LatB treatments used affect nuclear actin polymerization in addition to cytoplasmic actin[65].

Cell imaging and processing in time-lapse experiments. For MSD analyses, cells stably expressing GFP–Mdc1 and mCherry–HP1a were imaged before and after IR as described[24]. In brief, imaging after IR was done every 40 s for 60 min starting from 3–5 min after IR[7,24]. Ten *z* stacks at 0.8 μm distance were imaged for 0.005 ms for GFP, and 0.015 ms for mCherry. The Coolsnap HQ2 camera was set at 2×2 binning to maximize the light collected while minimizing light exposure and cell phototoxicity. Accordingly, these imaging conditions did not induce DSB formation or affect cell division (Extended Data Fig. 7a). All videos were corrected to compensate for modest photobleaching effects using softWoRx (Applied Precision/GE Healthcare). For each nucleus, 6–12 of the most stationary and spatially distant foci were tracked with Imaris (Bitplane) and used as a reference for registering the nucleus as previously described[24]. Foci were tracked in 3D using a semi-automated method and manually corrected to ensure optimal connections between time points[24]. For live imaging experiments, IR dose was lowered from 5 Gy to 1.7 Gy to reduce the average number of Mdc1 foci per nucleus and the frequencies of ambiguous tracks. Focus positional data were extracted in Excel and analysed in Matlab (MathWorks) using a customized script[24]. MSDs were calculated as previously described[24]. Positional data were also analysed using a customized script in R Studio to detect LDMs[24]. In brief, the script derives MSD curves for time intervals $\Delta t > 10$ and increasing starting time points for each trajectory, and identifies time intervals characterized by MSD graphs with increasing slopes[22,59]. LDMs shown in Fig. 4c and Extended Data Fig. 7d correspond to the largest contiguous time interval containing MSDs of increasing slope for each focus. The ability of the script to correctly detect directed motions was confirmed by independent MSD analysis of the positional data within the time intervals of the LDMs, as shown in Extended Data Fig. 7e. For live imaging of nuclear actin filaments, a stable cell line expressing F-actCB–GFP–NLS was exposed to 5 Gy IR. The same field of cells was imaged before and every 5 min after IR for 60 min. Ten *z* stacks at 0.8 μm distance were imaged, starting 3–12 min after IR. 3D volume reconstructions and movie generation were done in Imaris (Bitplane). Quantification of filament length and angles of actin branches was done using the Measurement Tool in Imaris.

IR sensitivity assay. The IR sensitivity assay for RNAi depletion experiments (Fig. 5c, left) was done as previously described[7]. Quantification of IR sensitivity for cells expressing actin(WT) or actin(R62D) (Fig. 5c, right) was done using a trypan blue exclusion test for cell viability[66].

Micronucleus assays. The micronucleus assay in Kc cells was done as previously described[7]. Micronuclei in NIH3T3 cells were quantified in cells fixed 72 h after IR based on DAPI and H3K9me3 staining.

Antibodies. Primary antibodies used in *Drosophila* cells were: anti-actin (1:1,000, Abcam, ab8224); anti-γH2Av (1:1,000, Rockland, 600-401-914); anti-Rad51 (1:1,000, gift from J. Kadonaga); anti-HA (1:1,000, Abcam, ab9134 for western blot; 1:1,000; Covance, 16B12 for immunofluorescence); anti-Flag (1:1,000, Sigma, F1804); anti-GFP (1:1,000, Invitrogen, AP11122 for western blot; 1:1,000 Aves Laboratory, GFP-1020 for immunofluorescence; Rockland, 600-101-215 for immunoprecipitation); anti-H3K9me2 (1:750, Wako Chemicals, MABI0307, 302-32369); anti-Wash (1:10, Developmental Studies Hybridoma Bank, P3H3[67]); anti-Wasp (1:5, Developmental Studies Hybridoma Bank, P5E1[68]); anti-Scar (1:10, Developmental Studies Hybridoma Bank, P1C1[68]); anti-Whamy (1:10, Developmental Studies Hybridoma Bank, P1D1[68]); anti-Arpc3A (1:10,000, for western blotting, gift from L. Cooley), anti-Arpc3B (1:500, for immunofluorescence, gift from L. Cooley), anti-MyoV (head) (1:500, gift from A. Eprussi), anti-Unc45 (1:500, gift from S. Bernstein), anti-Smc5 (SDI, 1:800, gift from G. Karpen[5]); anti-Smc6; (SDI, 1:800, gift from G. Karpen[5]); anti-Nup62 (1:1,000, gift from H. Ohkura); anti-Rad50 (1:1,000, gift from M. Gatti), anti-dPIAS (1:1,000, gift from G. Karpen), anti-TopBP1 (1:1,000, gift from M. Michael[69]); anti-HP1a (1:500, Developmental Studies Hybridoma Bank, C1A9[70]). Primary antibodies used in NIH3T3 cells were: anti-H3K9me3 (1:2,000, Abcam, ab8898); anti pH3S10 (1:4,000, Milipore, 06-570); anti-γH2AX (phospho-S139; 1:2,000, Abcam, ab26350). Secondary antibodies for immunofluorescence were from Life Technologies and Jackson Immunoresearch. Those used for western blotting were from Pierce and Santa Cruz Biotech. Antibodies were previously validated[5,7,44] or validated by comparing western blot or immunofluorescence signals in the presence of the protein of interest with signals after RNAi depletions, or immunofluorescence signals in the absence of primary antibodies.

Fly stocks and crosses. *Drosophila* were maintained on standard medium at 25 °C, which was prepared as previously described[71]. Stocks were obtained from BDSC (http://fly.bio.indiana.edu) or VDRC (https://stockcenter.vdrc.at/control/main) and are: Myo1A (BDSC #33971) y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8] = TRiP.HMS00298}attP2; Myo1B (BDSC #41689) y[1] v[1]; P{y[+t7.7] v[+t1.8] = TRiP.HMS02253}attP2; Arp3 (BDSC #32921) y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8] = TRiP.HMS00711}attP2; Wash (BDSC #62866) y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8] = TRiP.HMC05339}attP40; Scar (BDSC #31126) y[1] v[1];P{y[+t7.7] v[+t1.8] = TRiP.JF01599}attP2; Act5c-GAL4 (BDSC #4414) y[1] w[*]; P{w[+mC] = Act5C-GAL4}25FO1/CyO, y[+]; Unc45 (VDRC #v108868) P{KK101311}VIE-260B. The Smc5 trans-heterozygous mutant in Fig. 5f and

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 8i was *smc5*$^{7/19}$, previously described[7]. The wild-type control was *w*^{1118}. To obtain third instar larvae for karyotyping of neuroblast metaphase spread, RNAi lines were crossed to the Act5c–GAL4 line (balanced with CyO–GFP) and non-GFP larvae were picked for karyotyping as previously described[7].

Statistics and reproducibility. All statistical analyses were performed using Prism 6 software (Graphpad), using the statistical tests indicated in the individual figure legends. Detailed information about sample sizes across different replicates and *P* values are provided as Source Data. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

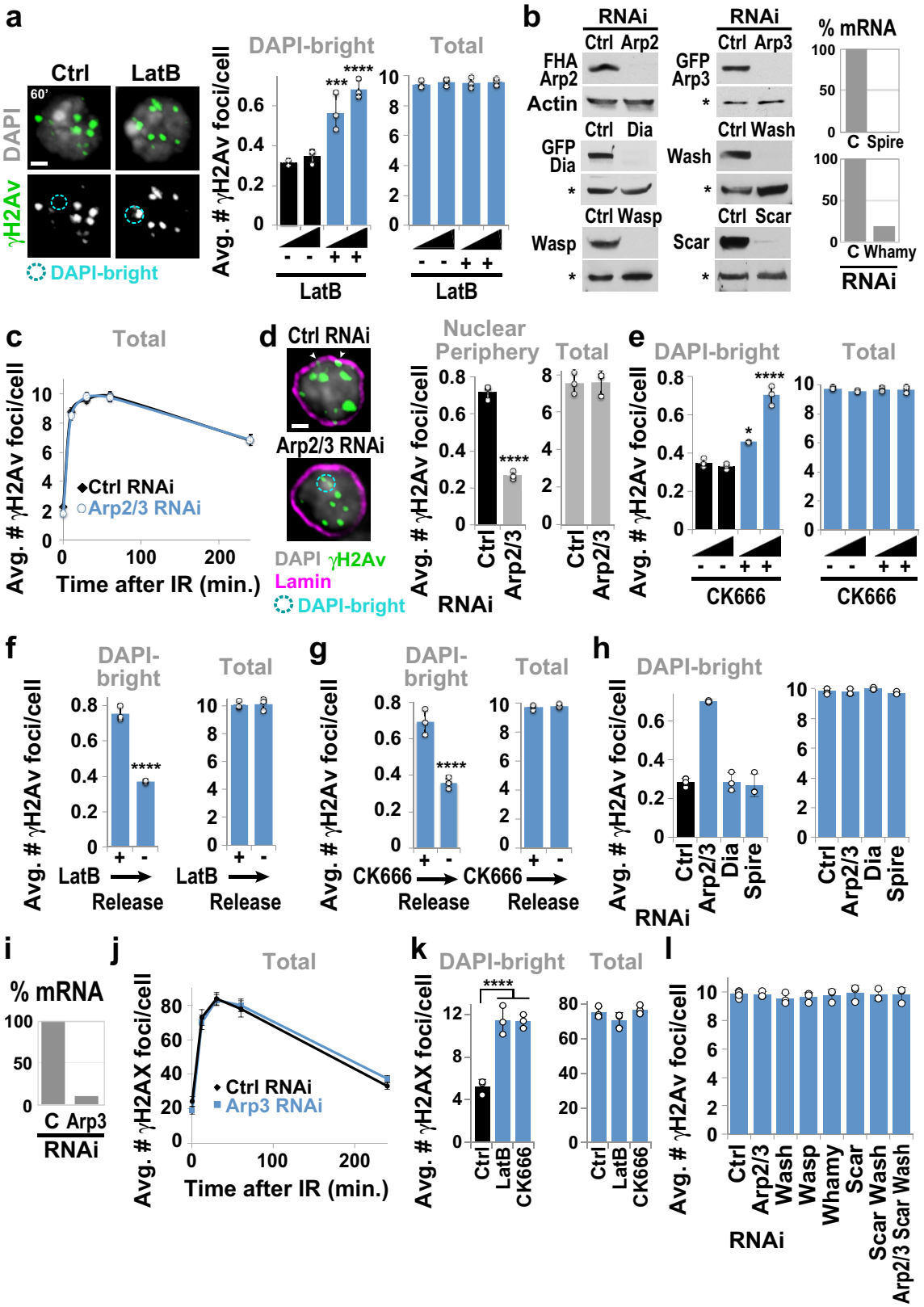
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. Custom scripts written in Matlab and R for image analysis have been published[24].

Data availability. All relevant data are included in the main manuscript and figures, Extended Data and Supplementary Information. Additional data are available from the corresponding author upon reasonable request.

43. Cherbas, L. & Gong, L. Cell lines. *Methods* **68**, 74–81 (2014).
44. Goodarzi, A. A., Noon, A. T. & Jeggo, P. A. The impact of heterochromatin on DSB repair. *Biochem. Soc. Trans.* **37**, 569–576 (2009).
45. Noon, A. T. et al. 53BP1-dependent robust localized KAP-1 phosphorylation is essential for heterochromatic DNA double-strand break repair. *Nat. Cell Biol.* **12**, 177–184 (2010).
46. Erhardt, S. et al. Genome-wide analysis reveals a cell cycle-dependent mechanism controlling centromere propagation. *J. Cell Biol.* **183**, 805–818 (2008).
47. Hudson, A. M. & Cooley, L. A subset of dynamic actin rearrangements in *Drosophila* requires the Arp2/3 complex. *J. Cell Biol.* **156**, 677–687 (2002).
48. Lee, C. F., Melkani, G. C. & Bernstein, S. I. The UNC-45 myosin chaperone: from worms to flies to vertebrates. *Int. Rev. Cell Mol. Biol.* **313**, 103–144 (2014).
49. Kokai, E. et al. Analysis of nuclear actin by overexpression of wild-type and actin mutant proteins. *Histochem. Cell Biol.* **141**, 123–135 (2014).
50. Plessner, M., Melak, M., Chinchilla, P., Baarlink, C. & Grosse, R. Nuclear F-actin formation and reorganization upon cell spreading. *J. Biol. Chem.* **290**, 11209–11216 (2015).
51. Krauss, J., López de Quinto, S., Nüsslein-Volhard, C. & Ephrussi, A. Myosin-V regulates oskar mRNA localization in the *Drosophila* oocyte. *Curr. Biol.* **19**, 1058–1063 (2009).
52. Cheeseman, I. M. & Desai, A. A combined approach for the localization and tandem affinity purification of protein complexes from metazoans. *Sci. STKE* **2005**, pl1 (2005).
53. Verboon, J. M., Sugumar, B. & Parkhurst, S. M. Wiskott-Aldrich syndrome proteins in the nucleus: aWASH with possibilities. *Nucleus* **6**, 349–359 (2015).
54. Verboon, J. M. et al. Wash interacts with lamin and affects global nuclear organization. *Curr. Biol.* **25**, 804–810 (2015).
55. Larracuente, A. M. & Ferree, P. M. Simple method for fluorescence DNA *in situ* hybridization to squashed chromosomes. *J. Vis. Exp.* **95**, 52288 (2015).
56. Small, J. V., Herzog, M. & Anderson, K. Actin filament organization in the fish keratocyte lamellipodium. *J. Cell Biol.* **129**, 1275–1286 (1995).
57. Sawasdichai, A., Chen, H. T., Abdul Hamid, N., Jayaraman, P. S. & Gaston, K. In situ subcellular fractionation of adherent and non-adherent mammalian cells. *J. Vis. Exp.* **41**, 1958 (2010).
58. Whelan, D. R. & Bell, T. D. Image artifacts in single molecule localization microscopy: why optimization of sample preparation protocols matters. *Sci. Rep.* **5**, 7924 (2015).
59. Pierobon, P. et al. Velocity, processivity, and individual steps of single myosin V molecules in live cells. *Biophys. J.* **96**, 4268–4275 (2009).
60. Leyton Puig, D. et al. Tips and tricks for artifact-free PFA-based fixation of the actin cytoskeleton and its regulatory proteins for single molecule localization super-resolution microscopy. *Protoc. Exch.* (2016).
61. Baarlink, C., Wang, H. & Grosse, R. Nuclear actin network assembly by formins regulates the SRF coactivator MAL. *Science* **340**, 864–867 (2013).
62. Yarrow, J. C., Lechler, T., Li, R. & Mitchison, T. J. Rapid de-localization of actin leading edge components with BDM treatment. *BMC Cell Biol.* **4**, 5 (2003).
63. Islam, K. et al. A myosin V inhibitor based on privileged chemical scaffolds. *Angew. Chem.* **49**, 8484–8488 (2010).
64. Nolen, B. J. et al. Characterization of two classes of small molecule inhibitors of Arp2/3 complex. *Nature* **460**, 1031–1034 (2009).
65. Belin, B. J., Cimini, B. A., Blackburn, E. H. & Mullins, R. D. Visualization of actin filaments and monomers in somatic cell nuclei. *Mol. Biol. Cell* **24**, 982–994 (2013).
66. Strober, W. Trypan blue exclusion test of cell viability. *Curr. Protoc. Immunol.* **111**, A3.B.1–A3.B.3 (2015).
67. Liu, R. et al. Wash functions downstream of Rho and links linear and branched actin nucleation factors. *Development* **136**, 2849–2860 (2009).
68. Rodriguez-Mesa, E., Abreu-Blanco, M. T., Rosales-Nieves, A. E. & Parkhurst, S. M. Developmental expression of *Drosophila* Wiskott–Aldrich Syndrome family proteins. *Dev Dyn.* **241**, 608–626 (2012).
69. Van Hatten, R. A. et al. The *Xenopus* Xmus101 protein is required for the recruitment of Cdc45 to origins of DNA replication. *J. Cell Biol.* **159**, 541–547 (2002).
70. James, T. C. & Elgin, S. C. Identification of a nonhistone chromosomal protein associated with heterochromatin in *Drosophila melanogaster* and its gene. *Mol. Cell. Biol.* **6**, 3862–3872 (1986).
71. Ren, C., Webster, P., Finkel, S. E. & Tower, J. Increased internal and external bacterial load during *Drosophila* aging without life-span trade-off. *Cell Metab.* **6**, 144–152 (2007).
72. Quinlan, M. E., Heuser, J. E., Kerkhoff, E. & Mullins, R. D. *Drosophila* Spire is an actin nucleation factor. *Nature* **433**, 382–388 (2005).
73. Afshar, K., Stuart, B. & Wasserman, S. A. Functional analysis of the *Drosophila* diaphanous FH protein in early embryonic development. *Development* **127**, 1887–1897 (2000).
74. Melak, M., Plessner, M. & Grosse, R. Actin visualization at a glance. *J. Cell Sci.* **130**, 525–530 (2017).
75. Schoenenberger, C. A. et al. Conformation-specific antibodies reveal distinct actin structures in the nucleus and the cytoplasm. *J. Struct. Biol.* **152**, 157–168 (2005).
76. Dopie, J., Skarp, K. P., Rajakylä, E. K., Tanhuanpää, K. & Vartiainen, M. K. Active maintenance of nuclear actin by importin 9 supports transcription. *Proc. Natl Acad. Sci. USA* **109**, E544–E552 (2012).
77. Senaratne, T. N., Joyce, E. F., Nguyen, S. C. & Wu, C. T. Investigating the interplay between sister chromatid cohesion and homolog pairing in *Drosophila* nuclei. *PLoS Genet.* **12**, e1006169 (2016).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

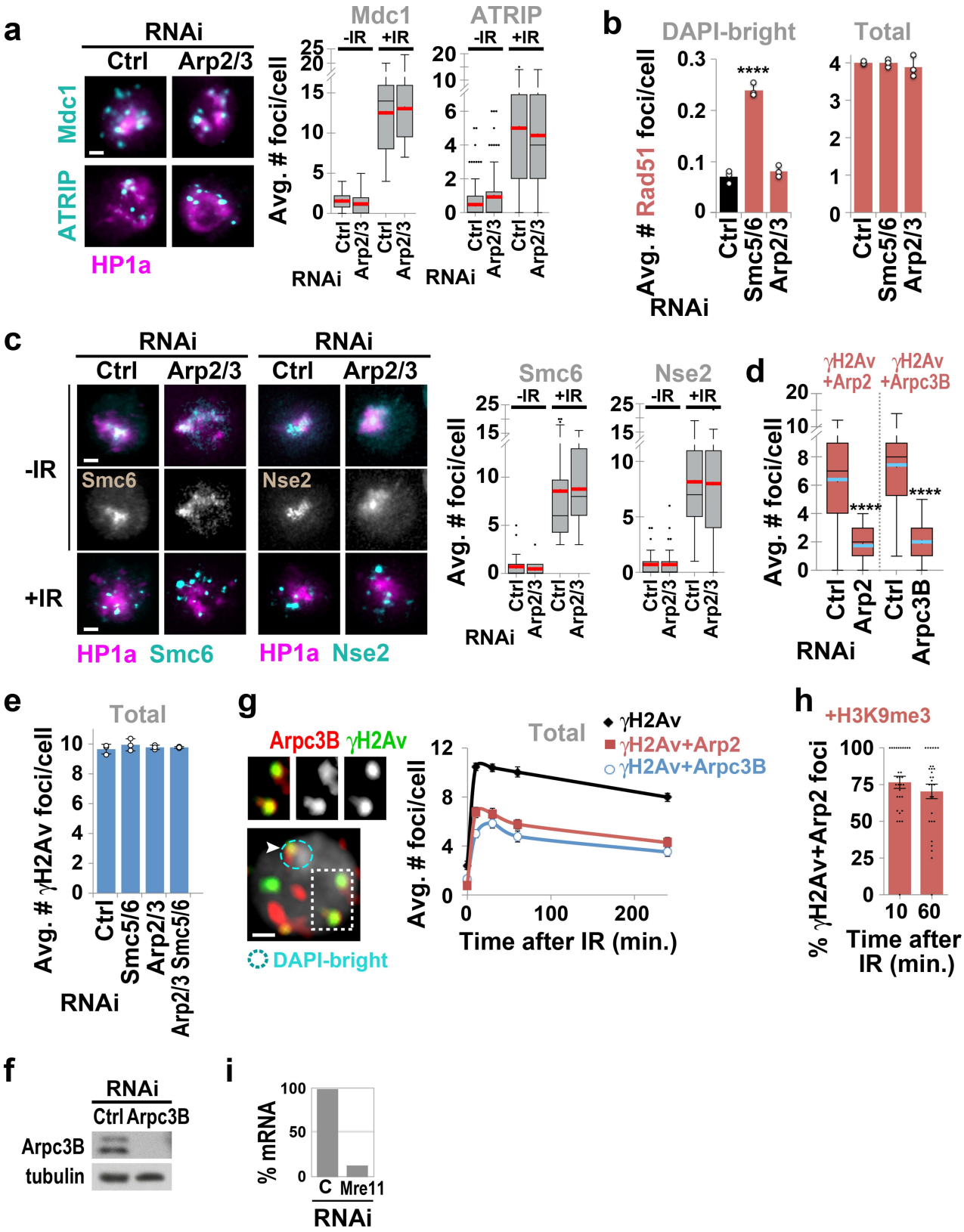


Extended Data Fig. 1 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 1 | Actin nucleators mediate relocalization of heterochromatic DSBs. a, Immunofluorescence and quantification of γH2Av foci in DAPI-bright heterochromatin (dashed circle), or total focus number, of Kc cells fixed 60 min after IR after treatment with LatB (+) or control (Ctrl, −). $^{****}P < 0.0001$, $^{***}P = 0.0008$ versus control, $n > 300$ cells per treatment. **b**, Western blotting or qPCR analyses show RNAi depletion efficiencies for indicated proteins in Kc cells. Actin or a background band (*) were used as loading controls. **c**, Quantification of total γH2Av foci from Fig. 1a. **d**, Immunofluorescence and quantification of γH2Av foci at the nuclear periphery (Lamin), or total focus number, 60 min after IR after indicated RNAi depletions in Kc cells. Only middle *z* stacks were used for quantifications, as previously described[7]. $^{****}P < 0.0001$, $n > 200$ cells per RNAi. **e**, As in **a** but for cells treated with CK666. $^{****}P < 0.0001$, $^{*}P = 0.0293$, $n > 300$ cells per treatment. **f**, **g**, Quantification of γH2Av foci in DAPI-bright heterochromatin or total focus number, after 60 min treatment with, and release from, LatB or CK666. $^{****}P < 0.0001$, $n > 200$ cells per condition. **h**, Quantification of γH2Av foci in DAPI-bright heterochromatin, or total focus number, in Kc cells fixed 60 min after IR, after RNAi depletion of actin nucleators Spire[72] or the formin Dia[73]. $^{****}P < 0.0001$ versus control, $n > 300$ cells per RNAi. **i**, qPCR analysis shows ARP3 RNAi depletion efficiency in NIH3T3 cells. **j**, Quantification of total γH2AX foci for Fig. 1b. **k**, Quantification of γH2AX foci in DAPI-bright heterochromatin, or total focus number, in NIH3T3 G2 cells treated with indicated chemicals, and fixed 60 min after IR. $^{****}P < 0.0001$ versus control, $n > 37$ cells per treatment. **l**, Quantification of total γH2Av foci for Fig. 1c. Scale bars, 1 μm. Data shown as mean ± s.e.m. in **c**, **j** and mean ± s.d. of at least three independent experiments in **a**, **d**–**h**, **k**, **l**. *P* values calculated using two-tailed Mann–Whitney test.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

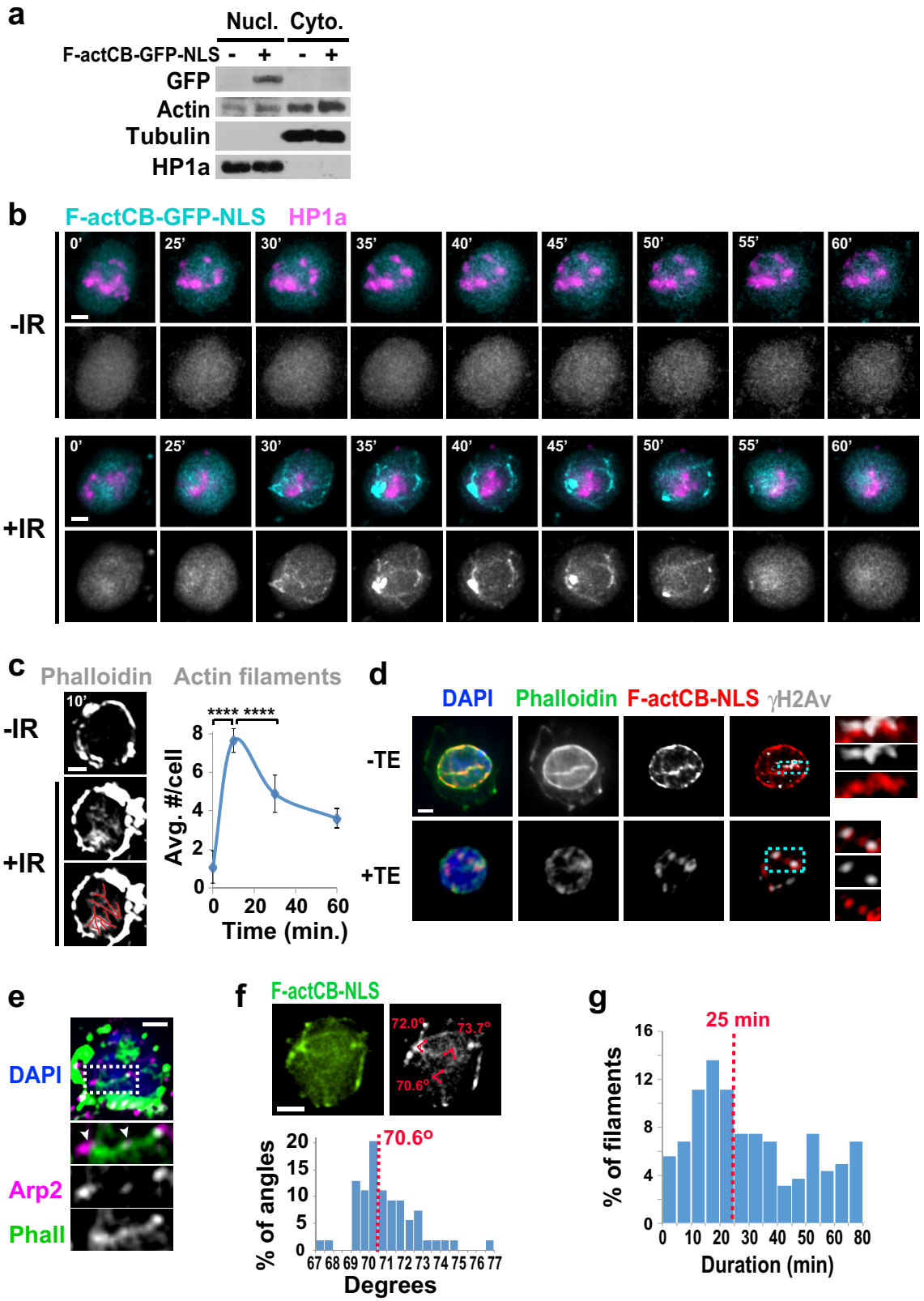


Extended Data Fig. 2 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 2 | Actin nucleators mediate relocalization of heterochromatic DSBs. a, Images 10 min after IR and quantification before (−) and 10 min after (+) IR of Mdc1 and Atrip foci in cells expressing GFP–Mdc1 or GFP–Atrip and mCherry–HP1a, after indicated RNAi depletions. $n \geq 48$ cells for Atrip; $n \geq 21$ cells for Mdc1. **b**, Quantification of Rad51 foci in cells fixed 60 min after IR and processed for immunofluorescence with anti-Rad51 antibodies after indicated RNAi depletions. Average number of foci shown relative to DAPI-bright or total foci. Smc5/6 RNAi results in abnormal formation of Rad51 foci in heterochromatin[5,7,8], and is used as positive control. $^{****}P < 0.0001$ versus control, unpaired *t*-test with Welch's correction, $n > 300$ cells per RNAi. **c**, Images and quantifications of Smc6 and Nse2 colocalizing with the HP1a domain before (−) IR and forming foci before and 15 min after (+) IR in cells expressing GFP–Smc6 or GFP–Nse2 and mCherry–HP1a after indicated RNAi depletions. $n = 35$ cells for Nse2 and $n \geq 23$ cells for Smc6. **d**, Quantification 30 min after IR of γH2Av foci colocalizing with FHA-tagged Arp2 or with Arpc3B after indicated RNAi depletions. $^{****}P < 0.0001$, two-tailed Mann–Whitney test, $n \geq 31$ cells per RNAi. **e**, Quantification of total foci for Fig. 1d. **f**, Western blot analyses show RNAi depletion efficiency for Arpc3B, and specificity of anti-Arp3cB antibodies. Tubulin is used as loading control. **g**, Immunofluorescence of γH2Av foci colocalizing with Arpc3B and quantification of total focus number for Fig. 1f. **h**, Quantification at indicated times after IR of the percentage of γH2Av foci colocalizing with FHA–Arp2 relative to H3K9me3 $n = 30$. **i**, qPCR analysis shows Mre11 RNAi efficiency. Scale bars, 1 μm. Data shown as mean ± s.e.m. in **a**, **c**, **d**, **g**, **h** and mean ± s.d. of at least three independent experiments in **b**, **e**.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

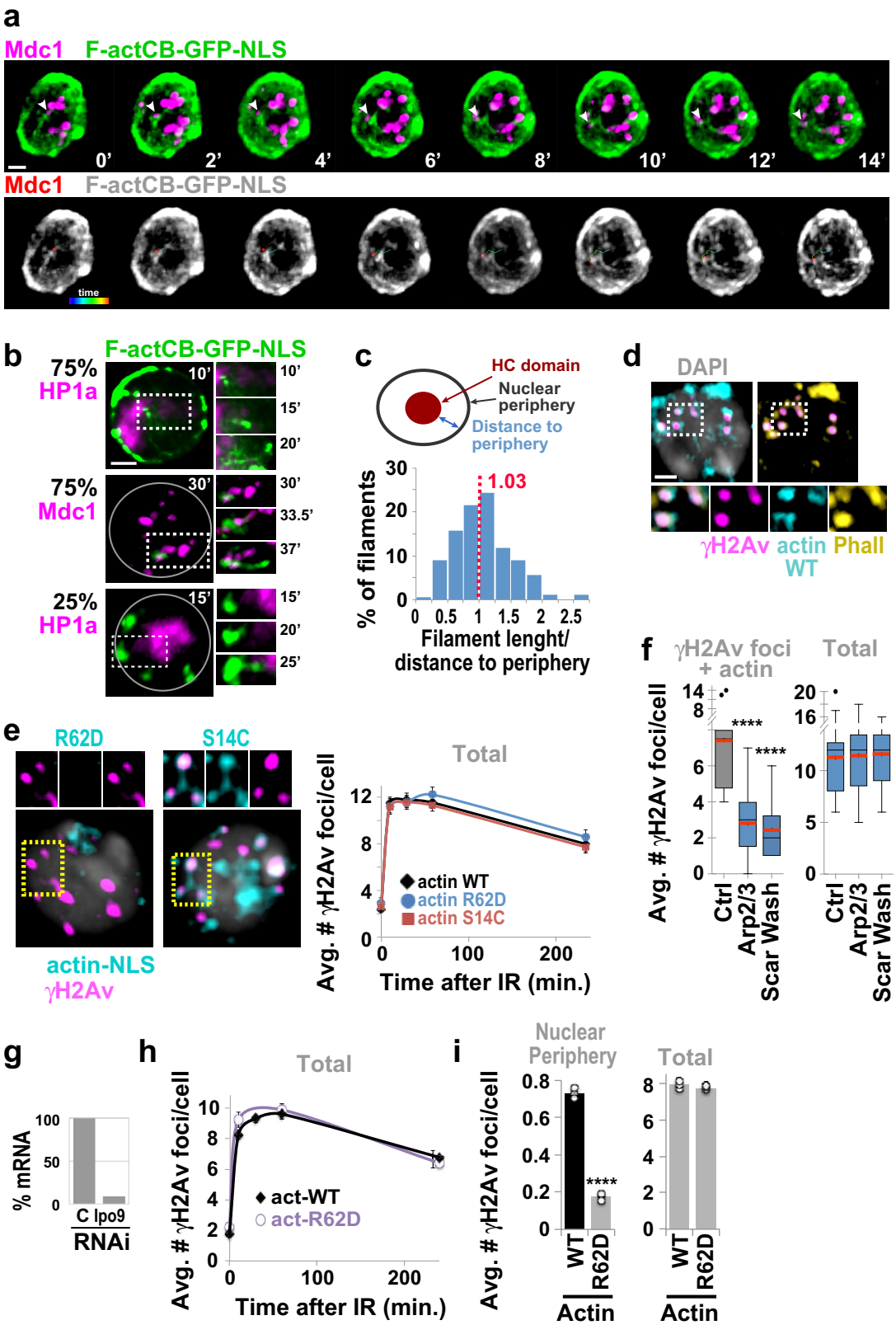


Extended Data Fig. 3 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 3 | Nuclear F-actin is required for relocalization. **a**, Western blot showing nuclear and cytoplasmic actin levels in cells stably transfected with F-actCB–GFP–NLS. HP1a and tubulin were used as loading controls and to check for purity of the fractions. Consistent with previous studies, expression of F-actCB–GFP–NLS does not alter nuclear actin levels[50,74], providing a non-invasive method for nuclear actin detection in vivo. **b**, Frames of time-lapse experiments (Fig. 2a, Supplementary Videos 1, 2) of cells expressing F-actCB–GFP–NLS and mCherry–HP1a signals treated (+) or not treated (−) with IR. Time points are from IR (+IR Video) or the start of imaging (−IR Video). **c**, Immunofluorescence and quantification of cells stained with phalloidin show the formation of canonical nuclear actin filaments[75] (highlighted in red) at indicated time points after IR. $****P < 0.0001$ for 10 min versus 0 min or 30 min, two-tailed Mann–Whitney test, $n \geq 51$ cells per time point. Error bars show s.d. of at least three independent experiments. **d**, Immunofluorescence and staining for GFP, phalloidin, and γH2Av, of cells expressing F-actCB–GFP–NLS, show examples of actin filaments coated with damage foci in the absence (−) of triton extraction (TE), and actin puncta colocalizing with γH2Av foci after (+) TE. Zoomed details highlight colocalizations. **e**, Immunofluorescence and staining for HA and phalloidin (Phall) of cells expressing FHA–Arp2 show enrichment of Arp2 along nuclear actin filaments, including at the base of actin branches (arrowheads). **f**, Measurement of F-actin branch angles in response to IR, in time-lapse experiments with cells expressing F-actCB–GFP–NLS. $n = 54$ angles. **g**, Quantification of the average duration of IR-induced actin filaments in time-lapse experiments with cells expressing F-actCB–GFP–NLS. $n = 162$ actin filaments. Dashed red lines indicate median values in **f**, **g**. Scale bars, 1 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

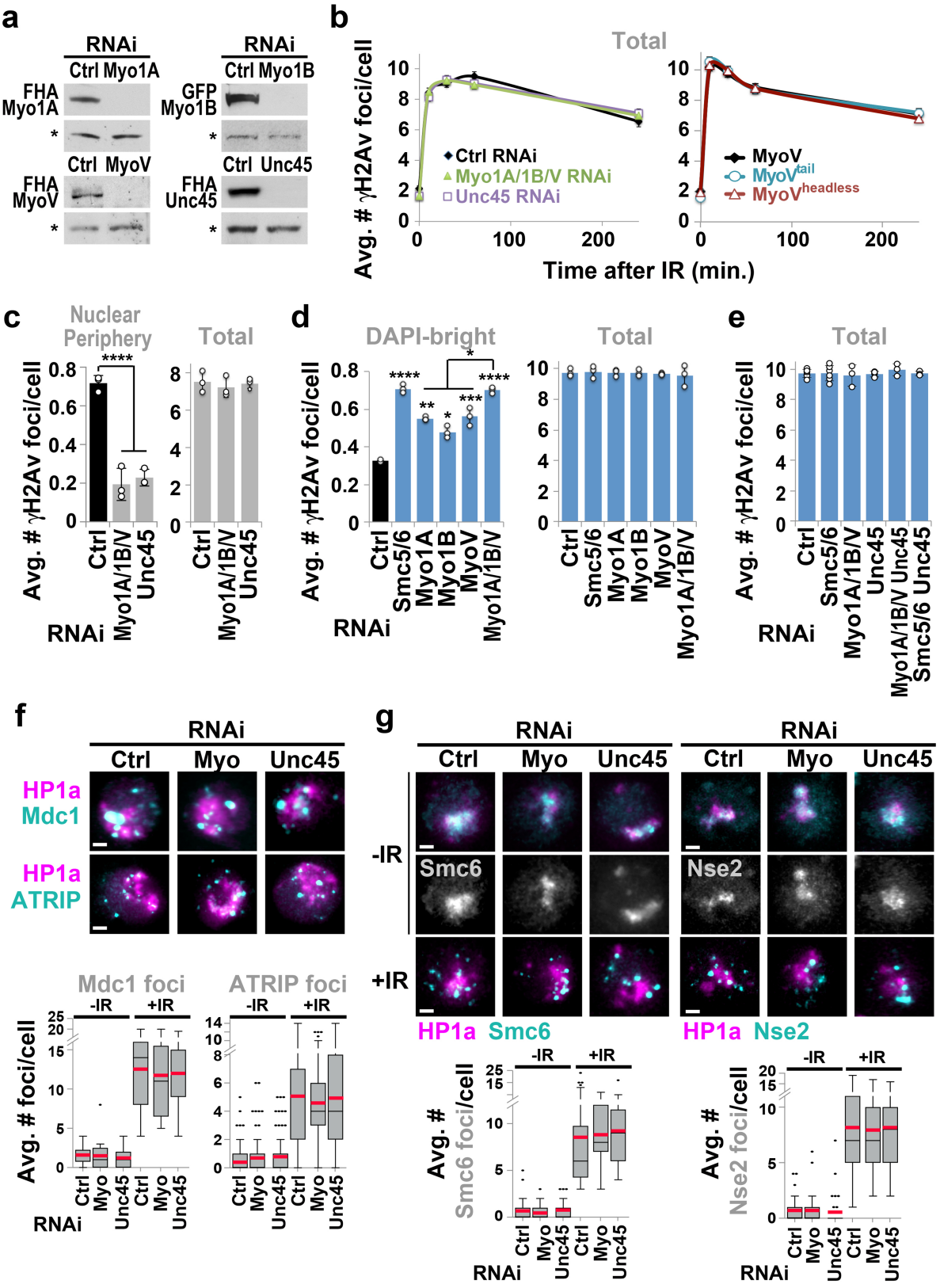


Extended Data Fig. 4 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 4 | Nuclear F-actin is required for relocalization. **a**, Frames from time-lapse experiment show cells expressing F-actCB–GFP–NLS and Mdc1–mCherry (Fig. 2c, Supplementary Video 3) with an example of a Mdc1 focus moving along an actin filament. Time indicates minutes from beginning of focus movement along the filament. Mdc1 focus tracking (bottom panels) was done in Imaris. **b**, Selected frames from cells expressing F-actCB–GFP–NLS and Mdc1–mCherry or HP1a–mCherry show examples of filament directionality at indicated time points after IR, that is, filaments emerging from the heterochromatin domain periphery (top), from repair foci (middle) or from the nuclear periphery (bottom). Percentages indicate the frequency with which each behaviour is observed. $n \geq 44$ filaments. **c**, Quantification of the length of actin filaments departing from the heterochromatin domain relative to the average distance between the HP1a domain periphery and the nuclear periphery. $n > 140$ filaments. Dashed red vertical line shows median value. **d**, Immunofluorescence and staining for Flag (actin(WT)), phalloidin (Phall), and γH2Av, of cells expressing Flag–NLS–actin(WT), show examples of actin signals in cells processed by triton extraction before fixation, 10 min after IR. Zoomed details highlight colocalizations. Scale bar, 1 μm. **e**, Immunofluorescence of cells expressing Flag–NLS-tagged versions of actin(R62D) and actin(S14C) shows colocalizations with γH2Av foci. Quantification shows total γH2Av foci for Fig. 2d. **f**, Quantification of γH2Av foci colocalizing with Flag–NLS–actin (WT) 10 min after IR, or total number of foci, after indicated RNAi depletions. $^{****}P < 0.0001$, $n \geq 14$ cells and $n \geq 158$ foci per RNAi. **g**, qPCR analysis shows Ipo9 RNAi depletion efficiency. Ipo9 depletion specifically blocks the transfer of actin monomers to the nucleus[14,76]. **h**, Quantification of total γH2Av foci for Fig. 2e. **i**, Quantification of cells fixed 60 min after IR shows γH2Av foci at the nuclear periphery (lamin), or total focus number, after expression of indicated actin forms. Only the middle *z* stack was used for quantifications. $^{****}P < 0.0001$, $n \geq 215$ cells per experiment. Error bars, s.e.m. in **e**, **f**, **h** and s.d. of three independent experiments in **i**. *P* values calculated with two-tailed Mann–Whitney test.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

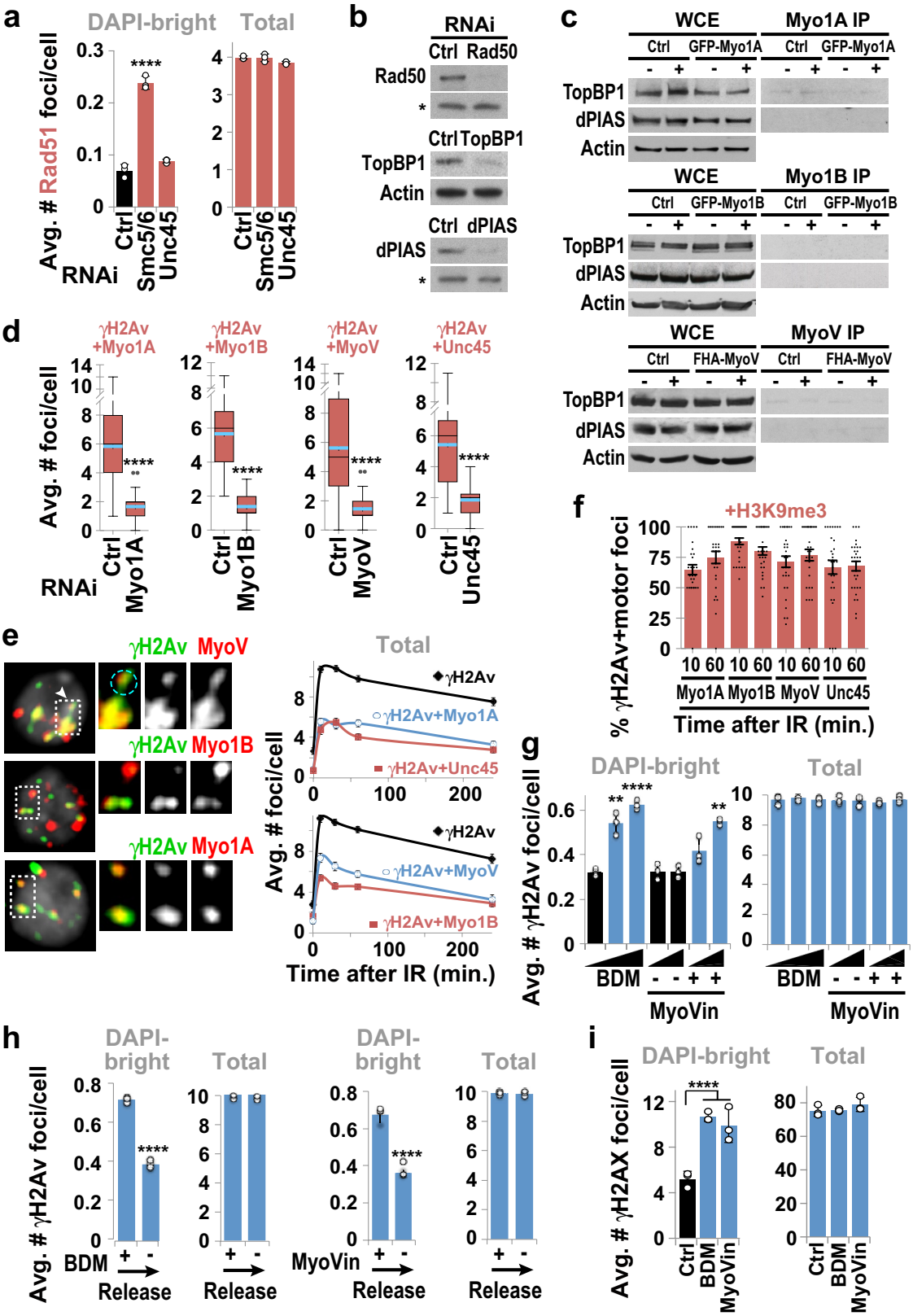


Extended Data Fig. 5 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 5 | Relocalization relies on nuclear myosins. **a**, Western blot analysis of cells expressing FHA-tagged components as indicated, shows RNAi depletion efficiency for indicated proteins. Background bands used as loading controls are indicated by an asterisk. **b**, Quantification shows total γH2Av foci for Fig. 3a. **c**, Quantification of γH2Av foci at the nuclear periphery (lamin) or total focus number in cells fixed 60 min after IR, after indicated RNAi depletions. $^{****}P < 0.0001$ versus control, $n \geq 200$ cells per RNAi. Only the middle Z-stack was used for quantifications. Control RNAi as in Extended Data Fig. 1d. **d**, Quantification of γH2Av foci in DAPI-bright heterochromatin, or total focus number, in cells fixed 60 min after IR, after indicated RNAi depletions. $^{****}P < 0.0001$, $^{***}P = 0.003$, $^{**}P = 0.0022$, $^{*}P \leq 0.0402$ versus control, $n \geq 200$ cells per RNAi. **e**, Quantification of total γH2Av foci for Fig. 3b. **f**, Images 10 min after IR and quantification at 0 (−IR) and 10 min after (+) IR of Mdc1 and Atrip foci in cells expressing GFP–Mdc1 or GFP–Atrip and mCherry–HP1a, after indicated RNAi depletions. $n \geq 75$ cells for Atrip and $n \geq 17$ cells for Mdc1. **g**, Images and quantifications of Smc6 and Nse2 colocalizing with the mCherry–HP1a domain before (−) IR and forming foci at 0 and 15 min after (+) IR in cells expressing GFP–Smc6 or GFP–Nse2 and mCherry–HP1a after indicated RNAi depletions. $n \geq 31$ cells for Nse2 and $n \geq 21$ cells for Smc6. In **f**, **g**, Myo indicates RNAi depletion of Myo1A, Myo1B and MyoV. Scale bars, 1 μm. Data shown as mean ± s.e.m. in **b**, **f**, **g** and mean ± s.d. of at least three independent experiments in **c**–**e**. All *P* values calculated with two-tailed Mann–Whitney test except one-tailed Mann–Whitney test was used for Myo1B RNAi in **d**.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

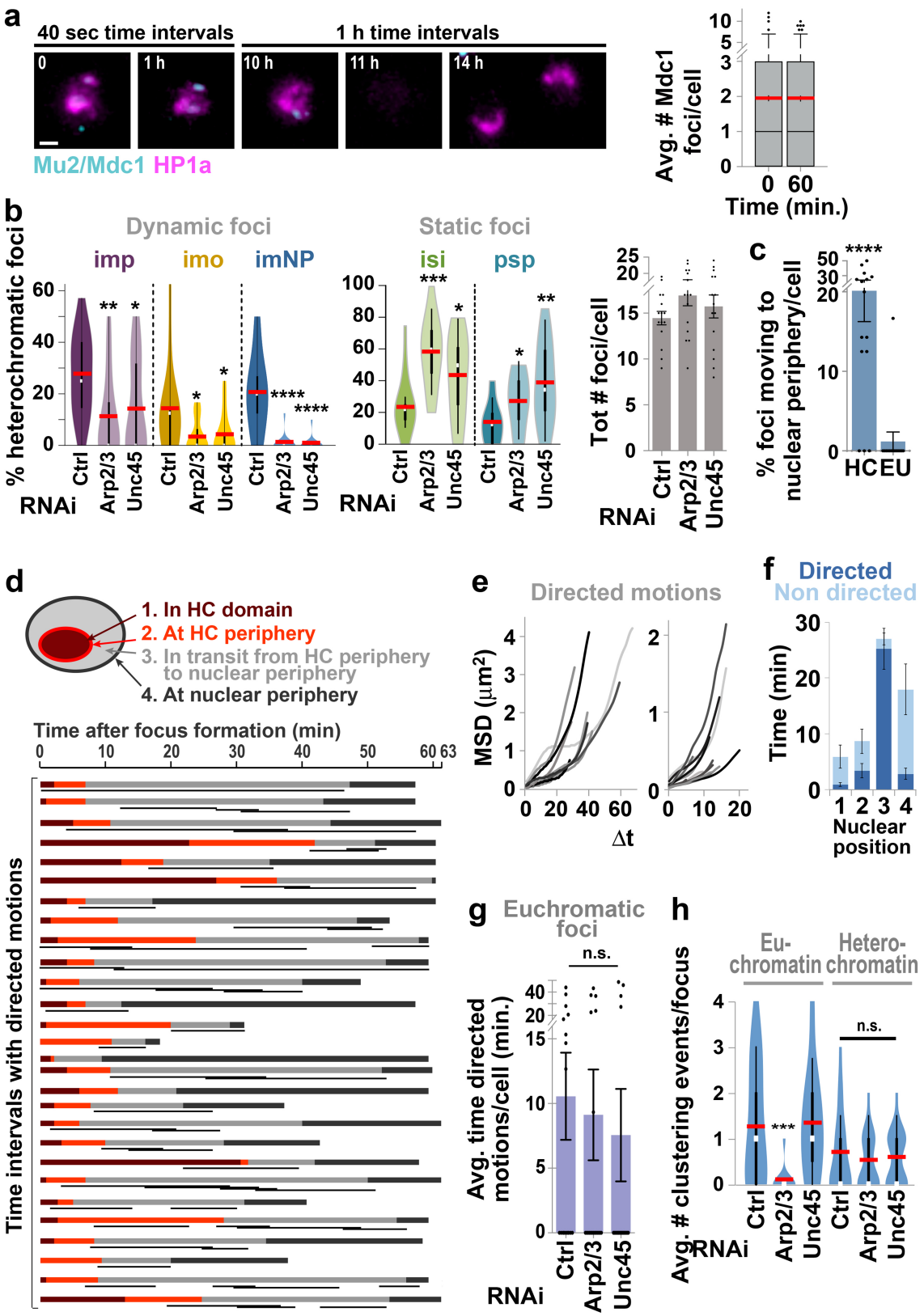


Extended Data Fig. 6 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 6 | Relocalization relies on nuclear myosins. **a**, Quantification of Rad51 foci in cells fixed 60 min after IR after indicated RNAi depletions shows the average number of foci relative to DAPI-bright heterochromatin or total foci. ****$P = 0.0001$ versus control, $n > 300$ cells per RNAi. Control and Smc5/6 RNAi as in Extended Data Fig. 2b. **b**, Western blot validation of the antibodies used in **c** and in Fig. 3c shows loss of Rad50, TopBP1 and dPIAS bands after corresponding RNAi depletions. **c**, Western blot analysis of the same immunoprecipitation shown in Fig. 3c probed with antibodies for dPIAS or TopBP1, with actin as the loading control. **d**, Quantification 30 min after IR of γH2Av foci colocalizing with FHA–Myo1A, GFP–Myo1B, MyoV or Unc45 after indicated RNAi depletions. ****$P < 0.0001$, $n \geq 21$ cells per RNAi. **e**, Immunofluorescence 10 min after IR and quantification at indicated time points after IR of γH2Av foci colocalizing with FHA–Myo1A, GFP–Myo1B, MyoV or Unc45. **f**, Quantification at indicated times after IR of the percentage of foci of γH2Av + FHA–Myo1A, GFP–Myo1B, MyoV or Unc45, colocalizing with H3K9me3. $n = 30$ cells per experiment per time point. **g**, Quantification of Kc cells fixed 60 min after IR shows γH2Av foci in DAPI-bright heterochromatin, or total focus number, after treatment with BDM, MyoVin or in controls. ****$P < 0.0001$, **$P < 0.008$, $n \geq 300$ cells per treatment. **h**, Quantification of γH2Av foci in DAPI-bright heterochromatin or total foci, after 60 min treatment with and release from BDM or MyoVin. ****$P < 0.0001$, $n \geq 190$ Kc cells per treatment. **i**, As in **g**, except NIH3T3 cells were used and γH2AX foci were quantified. Control RNAi as in Extended Data Fig. 1k. ****$P < 0.0001$, $n \geq 37$ cells and $n \geq 2{,}760$ foci per treatment. Data shown as mean ± s.e.m. in **d**–**f** and mean ± s.d. of at least three independent experiments in **a**, **g**–**i**. All P values calculated with two-tailed Mann–Whitney test.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

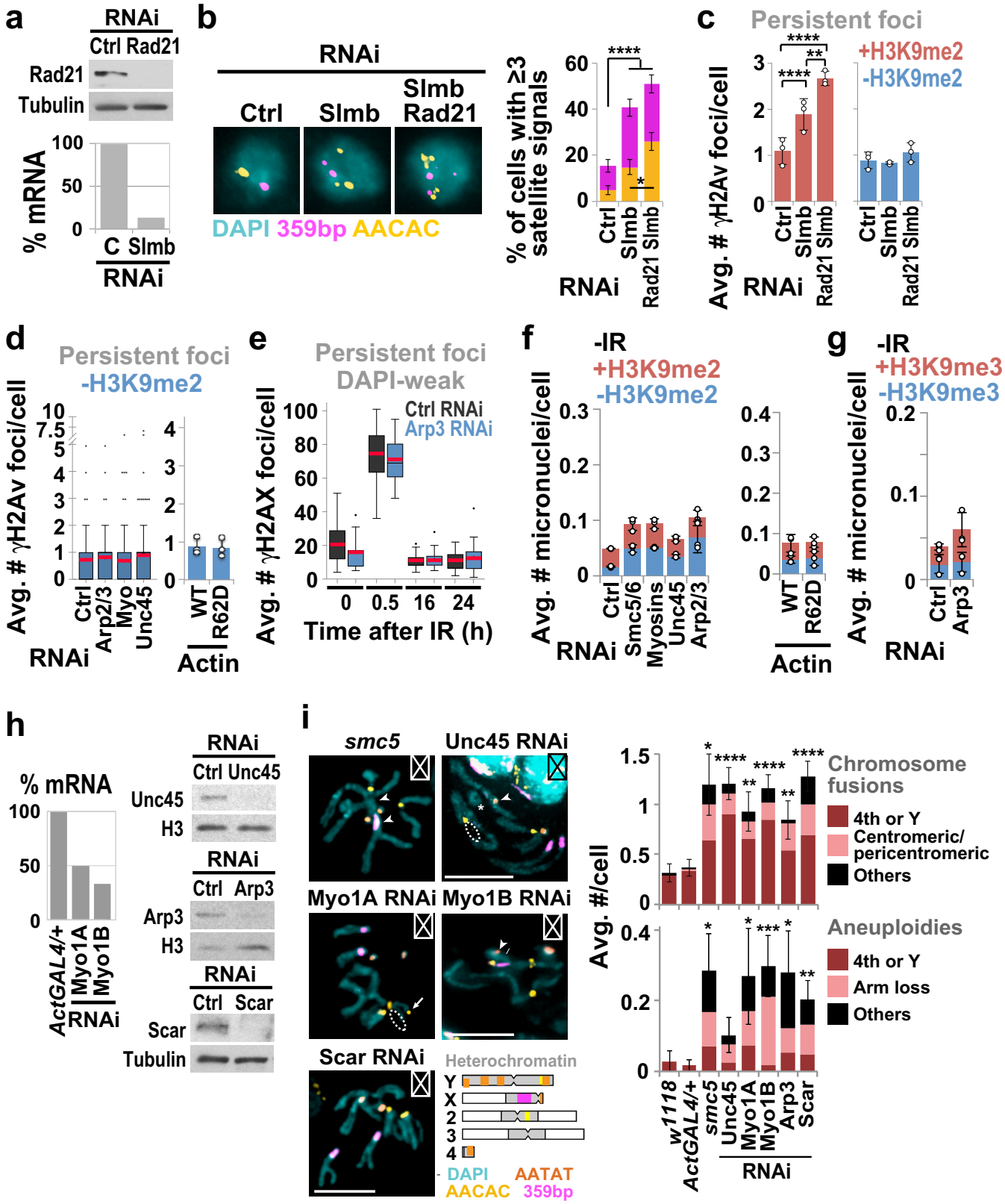


Extended Data Fig. 7 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 7 | Actin nucleators and myosin drive directed motions of heterochromatic DSBs. a, Selected images and quantification at indicated time points of Mdc1 foci in cells expressing GFP–Mdc1 and mCherry–HP1a show no effect of the imaging conditions used in Fig. 4 on the ability of cells to divide or the total number of repair foci. Only representative time points are shown. $n = 82$ cells. Scale bar, 1 μm. **b**, Left and middle, quantification and violin plot display of the frequency of heterochromatic Mdc1 foci moving relative to the HP1a domain during the time-lapse experiments shown in Fig. 4, after indicated RNAi depletions. White box, median; red line, mean; vertical black lines, values that occur 95% and 50% of the time. 'Dynamic HC foci' include foci moving from inside the HP1a domain to: the nuclear periphery (imNP); outside the HP1a domain (imo); or the periphery of the HP1a domain (imp). 'Static HC foci' include foci remaining inside (isi) or at the periphery of the HP1a domain (psp) throughout the 1-h timecourse. Colours reflect the categories analysed in Fig. 4b. Right, quantification of the average number of Mdc1 foci from Fig. 4a–d and **b**, after indicated RNAi depletions. $*P < 0.05$, $**P < 0.006$, $***P = 0.0002$; $****P < 0.0001$, two-tailed Mann–Whitney test, $n \geq 217$ foci per RNAi. **c**, Quantification of the percentage of heterochromatic (HC) and euchromatic (EU) Mdc1 foci moving to the nuclear periphery in time-lapse experiments from Fig. 4a. $****P < 0.0001$; two-tailed Mann–Whitney test, $n \geq 57$ foci. **d**, LDM analysis of Mdc1 foci that reach the nuclear periphery from Fig. 4b. Each coloured horizontal bar represents a focus and its duration in each nuclear compartment, as indicated. The black segments under each bar are LDMs for each focus. $n = 28$ foci. **e**, MSD curves generated from the positional data corresponding to the time points that contain LDMs in **d**. Longer (left) and shorter (right) LDMs are presented as independent graphs for clarity. **f**, Quantification of the analysis shown in Fig. 4c and **d** shows the average duration of directed motions (LDMs) or non-directed motions in the nuclear locations defined in **d**. $P < 0.0001$ for directed motions in 3 versus 1, 2, 4, unpaired *t*-test. **g**, Quantification of LDM durations for euchromatic foci shown in Fig. 4d, after indicated RNAi depletions. **h**, Quantification and violin plot of Mdc1 focus clustering events in time-lapse experiments for Fig. 4d. Graphical display as in **b**. An example of clustering is highlighted in Supplementary Video 4. $***P = 0.0003$, two-tailed Mann–Whitney test, $n \geq 20$ cells per RNAi. Error bars, s.e.m in **a**, **b** (right graph), **c**, **f**, **g**.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 8 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 8 | Actin nucleators and myosins promote heterochromatin repair and stability. a, Western blotting and qPCR analysis show RNAi depletion efficiency for Rad21 and Slmb as indicated. Tubulin is a loading control. **b**, FISH analysis and quantification show the effect of indicated RNAi on the number of cells with ≥ 3 AACAC or 359bp satellites, reflecting disruption of homologous and/or sister pairing[77]. $^{**}P = 0.0472$; $^{****}P < 0.0001$, two-tailed Mann–Whitney test, $n > 120$ cells per RNAi. **c**, Quantification shows γH2Av foci associated (+) or not associated (−) with H3K9me2 signals 20 h after IR following indicated RNAi depletions. $^{**}P = 0.0015$; $^{****}P < 0.0001$, two-tailed Mann–Whitney test, $n > 420$ cells per RNAi. **d**, Quantification of Kc cells from the experiment in Fig. 5a shows γH2Av foci not associated (−) with H3K9me2 signals 20 h after IR and following indicated RNAi depletions. **e**, Quantification of NIH3T3 cells from the experiment in Fig. 5b shows γH2AX foci not associated (−) with DAPI-bright signals 16 and 24 h after IR and following indicated RNAi depletions. **f**, Quantification of micronuclei in non-irradiated Kc cells from Fig. 5d. **g**, Quantification of micronuclei in non-irradiated NIH3T3 cells from Fig. 5e. **h**, qPCR and western blotting analyses show RNAi depletion efficiency of indicated components in third instar larvae for **i** and Fig. 5f. **i**, Images and quantification of chromosomal aberrations in karyotypes from Fig. 5f. Images show extra satellites (arrows), chromosome fusions (arrowhead) and chromosome arm losses (dashed circles). The diagram of *Drosophila* chromosomes indicates the positions of the main satellites detected by FISH. Aneuploidies and chromosome fusions from Fig. 5f were further categorized to highlight rearrangements involving centromeric regions or chromosomes that are predominantly heterochromatic (fourth or Y). $^{****}P \leq 0.0001$, $^{**}P \leq 0.0079$, $^{*}P \leq 0.0433$, unpaired *t*-test with Welsh correction. Error bars, s.e.m. in **b**, **d** (RNAi), **e**, **i** and s.d. of at least three independent experiments in **c**, **d** (actin), **f**, **g**.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ARTICLE

https://doi.org/10.1038/s41586-018-0237-5

Nuclear ARP2/3 drives DNA break clustering for homology-directed repair

Benjamin R. Schrank[1], Tomas Aparicio[1], Yinyin Li[2], Wakam Chang[3], Brian T. Chait[2], Gregg G. Gundersen[3], Max E. Gottesman[4] & Jean Gautier[1,5]*

DNA double-strand breaks repaired by non-homologous end joining display limited DNA end-processing and chromosomal mobility. By contrast, double-strand breaks undergoing homology-directed repair exhibit extensive processing and enhanced motion. The molecular basis of this movement is unknown. Here, using *Xenopus laevis* cell-free extracts and mammalian cells, we establish that nuclear actin, WASP, and the actin-nucleating ARP2/3 complex are recruited to damaged chromatin undergoing homology-directed repair. We demonstrate that nuclear actin polymerization is required for the migration of a subset of double-strand breaks into discrete sub-nuclear clusters. Actin-driven movements specifically affect double-strand breaks repaired by homology-directed repair in G2 cell cycle phase; inhibition of actin nucleation impairs DNA end-processing and homology-directed repair. By contrast, ARP2/3 is not enriched at double-strand breaks repaired by non-homologous end joining and does not regulate non-homologous end joining. Our findings establish that nuclear actin-based mobility shapes chromatin organization by generating repair domains that are essential for homology-directed repair in eukaryotic cells.

DNA double-strand breaks (DSBs) induce chromatin movement. In budding yeast, which repair DSBs primarily by homology-directed repair (HDR), induction of a single chromosomal break triggers increased local mobility: the DSB mean-square displacement is substantially higher than that of an undamaged region[1,2]. Moreover, multiple DSBs form clusters after traversing long distances[3]. DSB clustering may facilitate homology search, increase repair efficiency or shield breaks from misrepair[4,5]. These movements are intricately related to HDR. Factors that are critical for initiation of resection and downstream recombination are essential for DSB mobility in yeast[1,2]. In mammalian cells, DSBs are often described as more stable, suggesting that non-homologous end joining (NHEJ), the predominant repair pathway, limits its movement[6–8]. However, in human HeLa cells, RAD51-positive DSBs induced by alpha particles form clusters[4]. Similarly, damaged telomeres in human U2OS cells that are maintained by recombination merge in a RAD51-dependent manner[9]. Moreover, damaged active genes cluster in preparation for HDR[5]. Movement of deprotected mouse telomeres requires the LINC (linker of nucleoskeleton and cytoskeleton) complex, which transmits cytoskeletal forces from the cytoplasm to the nucleus[10]. The molecular basis of DSB movement and its role in DNA repair remain unclear.

The machinery that drives actin polymerization in the cytoplasm is also found in the nucleus[11]. Specifically, the ARP2/3 complex and its activator WASP, a Wiskott–Aldrich syndrome (WAS) protein family member, are located in both cellular compartments[12–14]. WASP brings the ARP2 and ARP3 subunits into close proximity to activate the complex and enable filament elongation[15]. Genotoxic agents trigger actin polymerization in the nucleoplasm of mammalian cells[16]; however, the role of actin polymerization in DSB repair has not been characterized.

Actin nucleators bind damaged chromatin

We performed an unbiased proteomics screen to document the recruitment of proteins to chromosomal DSBs in cell-free S-phase extracts derived from *Xenopus* eggs. Peptides from control or DSB-containing chromatin protein fractions were labelled with isobaric tags and subjected to liquid chromatography–mass spectrometry. We observed enrichment of known DSB repair regulators and proteins not previously associated with the DNA damage response (Extended Data Fig. 1a). Among these proteins were all seven subunits of the ARP2/3 complex, as well as β-actin and capping proteins (Extended Data Fig. 1a). We confirmed by western blotting that β-actin, ARPC4, and CAPZβ are recruited to MRE11-enriched, DSB-containing chromatin (Fig. 1a). We next investigated whether actin enrichment at chromosomal DSBs required DNA damage signalling. Inhibition of the phosphatidylinositol-3-OH kinase (PI(3)K)-like kinases ATM and ATR reduced the binding of actin complexes to damaged chromatin (Extended Data Fig. 1b, c). Moreover, treatment with the small molecule inhibitor CK-666, which stabilizes the ARP2/3 complex in an open, inactive conformation[17,18], decreased enrichment of ARPC4, β-actin, and CAPZβ in damaged chromatin (Fig. 1a, b). Overall, these results reveal that PI(3)K-like kinases and the ARP2/3 complex regulate the assembly of polymerized actin at chromosomal DSBs in *Xenopus* extracts.

WASP and ARP2/3 bind DSBs undergoing HDR

We next tested whether WASP localized to DSB foci in mammalian cells. Generation of DSBs by neocarzinostatin (NCS), a radiomimetic antibiotic, induced the formation of WASP foci in U2OS cells (Fig. 2a, b). Moreover, WASP substantially co-localized with γH2AX, which marks large chromatin domains surrounding DSBs[19], suggesting that sites of DNA repair contain WASP (Fig. 2c). Similarly, WASP foci arose in mouse-tail fibroblasts (MTFs) after DSB generation and co-localized with γH2AX (Extended Data Fig. 1d–f).

We next investigated whether actin complexes are recruited in close proximity to DSBs. Accordingly, we performed chromatin immunoprecipitation (ChIP) experiments in U2OS cells in which genome-wide DSBs had been generated by nuclear translocation of the AsiSI

[1]Institute for Cancer Genetics, College of Physicians and Surgeons, Columbia University, New York, NY, USA. [2]Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY, USA. [3]Department of Pathology and Cell Biology, College of Physicians and Surgeons, Columbia University, New York, NY, USA. [4]Department of Biochemistry and Biophysics, College of Physicians and Surgeons, Columbia University, New York, NY, USA. [5]Department of Genetics and Development, College of Physicians and Surgeons, Columbia University, New York, NY, USA. *e-mail: jg130@cumc.columbia.edu

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

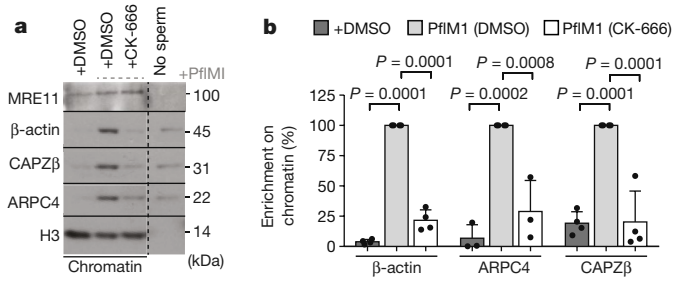


Fig. 1 | Actin complexes are recruited to damaged chromatin. a, Western blot showing enrichment of actin complexes in chromatin damaged using the PflMI restriction endonuclease (+PflMI). MRE11 indicates DNA damage. **b**, Protein quantification in chromatin relative to +PflMI samples. *P* calculated by one-way ANOVA with multiple comparisons; data shown as mean and s.d.; $n = 4$, 3, and 4 independent experiments, left to right.

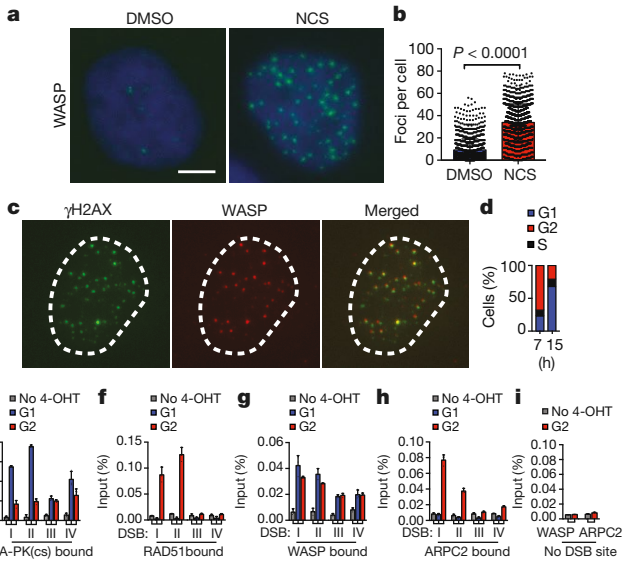


Fig. 2 | ARP2/3 and WASP co-localize at HDR breaks. a, Representative U2OS cells with WASP foci. **b**, Quantification of WASP foci. *P* calculated by two-sided Mann–Whitney test; data shown as mean and s.d.; $n = 1{,}231$ (DMSO), 1,327 nuclei (NCS). **c**, Co-localization of γH2AX with WASP ($n = 30$ nuclei; $r = 0.60 \pm 0.07$, Pearson). **d**, Cell cycle distribution of synchronized ER-AsiSI U2OS cells (U2OS cells stably expressing the AsiSI restriction endonuclease tethered to the oestrogen receptor) after thymidine release. **e**–**h**, ChIP showing enrichment of DNA-PKcs (**e**), RAD51 (**f**), WASP (**g**), and ARPC2 (**h**) at DSBs I–IV in G1 and G2 cells. Mean and s.d. ($n = 3$ technical replicates) of a representative experiment (out of two independent experiments). **i**, Mean and s.d. enrichment of ARPC2 and WASP at an undamaged site in G2 cells ($n = 3$ technical replicates) of a representative experiment (out of two independent experiments). Scale bar, 10 μm.

restriction enzyme[20]. We interrogated four DSB sites, annotated DSBs I–IV, in synchronized cells during the G1 or G2 phase of the cell cycle (Fig. 2d). First, we monitored the recruitment of RAD51 and the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs). RAD51 has been shown to localize to a subset of these DSBs[20]. DNA-PKcs was recruited to all four DSBs in both cell-cycle phases (Fig. 2e). By contrast, RAD51 bound exclusively to DSBs I and II, and this binding was more robust in the G2 phase than in the G1 phase (Fig. 2f). Together, these results demonstrate that HDR can occur at a subset of DSBs in G2 and distinguished from DSBs undergoing NHEJ, the more frequent mode of repair.

We next tested whether actin filament nucleators localize to DSB sites. WASP accumulated at DSBs I and II in both cell cycle phases (Fig. 2g), whereas ARPC2 recruitment was restricted to G2 (Fig. 2h). Notably, WASP was also enriched at DSBs III and IV, whereas ARPC2 was not (Fig. 2g, h). Neither WASP nor ARPC2 were enriched at an undamaged site (Fig. 2i). Collectively, these data establish that WASP binds DSBs independently of repair pathway choice but specifically activates ARP2/3 at DSBs undergoing HDR. This supports a model in which actin polymerization is primed at all DSBs but is triggered only during the assembly of HDR machinery.

ARP2/3 inactivation impairs DSB movement

The ARP2/3 complex generates propulsive forces by nucleating a highly-branched network of actin filaments. As WASP and ARP2/3 localized to RAD51-bound DSBs, we investigated whether their activity induced movement and clustering of HDR machinery. Accordingly, we visualized DSB movements in a U2OS cell line that stably expresses RAD52 tagged with mCherry and 53BP1 tagged with yellow fluorescent protein (YFP)[21]. RAD52 mediates recombination and forms foci exclusively in S/G2 cells[21,22]. Generation of DSBs in S/G2 cells yielded robust induction of RAD52–mCherry foci (Fig. 3a). Strikingly, RAD52 foci clustered and merged over a 60-min interval (Fig. 3a, b and Supplementary Video 1a–d). Coalescing foci increased in intensity and size and remained merged for a minimum of 15 min (Fig. 3b, c). Notably, inactivation of ARP2/3 using CK-666 decreased the frequency of clustering events and prevented the increase in the size of foci (Fig. 3c, d and Supplementary Video 1e, f). We subsequently compared DSB movements in the presence of CK-666 with those in the presence of CK-689, an inactive control analogue[18]. Quantitative analysis of DSB tracks in CK-689-treated cells indicated that RAD52 foci explored a mean cumulative distance of 4.1 μm over 100 min (Fig. 3e, f). Mean-square displacement (MSD) analysis of Rad52 foci revealed a diffusion coefficient $D(t)$ of $3.7 \times 10^{-5}\,\mu m^2\,s^{-1}$ (Fig. 3g and Extended Data Fig. 2a). Motion type analysis was consistent with confined Brownian motion. Critically, inhibition of ARP2/3 diminished the distance travelled by DSBs over the 100-min period (Fig. 3e, f). Furthermore, CK-666 treatment decreased the RAD52 diffusion coefficient from $3.7 \times 10^{-5}\,\mu m^2\,s^{-1}$ to $1.8 \times 10^{-5}\,\mu m^2\,s^{-1}$ (Fig. 3g).

We subsequently monitored DSBs bound by replication protein A (RPA) in MTFs expressing RPA32 tagged with enhanced GFP and a nuclear localization sequence (RPA32–pEGFP–NLS). Similar to the behaviour of RAD52–mCherry foci, RPA32 foci frequently merged over a 60-min interval (Fig. 3h and Supplementary Video 1g, h). Notably, CK-666 treatment diminished the distance travelled by RPA foci relative to CK-689 treatment (Fig. 3i). Moreover, the diffusion coefficient calculated for RPA foci in cells treated with CK-666 decreased from $4.1 \times 10^{-5}\,\mu m^2\,s^{-1}$ (for cells treated with CK-689) to $2.1 \times 10^{-5}\,\mu m^2\,s^{-1}$ (Extended Data Fig. 2a). Finally, we used MTFs harbouring a conditional *Arpc2* allele[23]. Upon 4-OHT treatment, these cells are depleted of ARPC2[23] (Fig. 3j). Consistent with CK-666-treated cells, RPA movements were significantly attenuated in ARCP2-depleted cells (Fig. 3i–l). CK-689 affected RAD52 and RPA foci movement only minimally relative to DMSO (Extended Data Fig. 2a, b).

The formation of Rad51 filaments requires single-stranded DNA (ssDNA)–RPA intermediates. Thus, we assessed how ARP2/3 inactivation affected the behaviour of RAD51 foci. To circumvent the lack of a functional fluorescent RAD51 reporter, we used Icy bioimaging software and its spatial analysis plug-in (http://www.icy.bioimage analysis.org) to assess clustering of RAD51 foci. This approach was validated in a study showing ATM-dependent clustering of γH2AX foci in fixed cells[24]. Analysis with Icy software revealed that RAD51 foci were highly clustered in the nucleus and that CK-666 reduced this clustering (Extended Data Fig. 3a–d). CK-666 treatment did not affect cell viability, cell cycle progression, nuclear size, nuclear morphology, or the expression of RPA or RAD51 (Extended Data Fig. 4a–g). These experiments demonstrate that ARP2/3-mediated actin polymerization enhances DSB motion during HDR, thereby increasing the clustering of foci.

During the G1 phase of the cell cycle, 53BP1 forms foci and regulates end-joining events. Consistent with the lack of ARP2/3 at DSBs

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

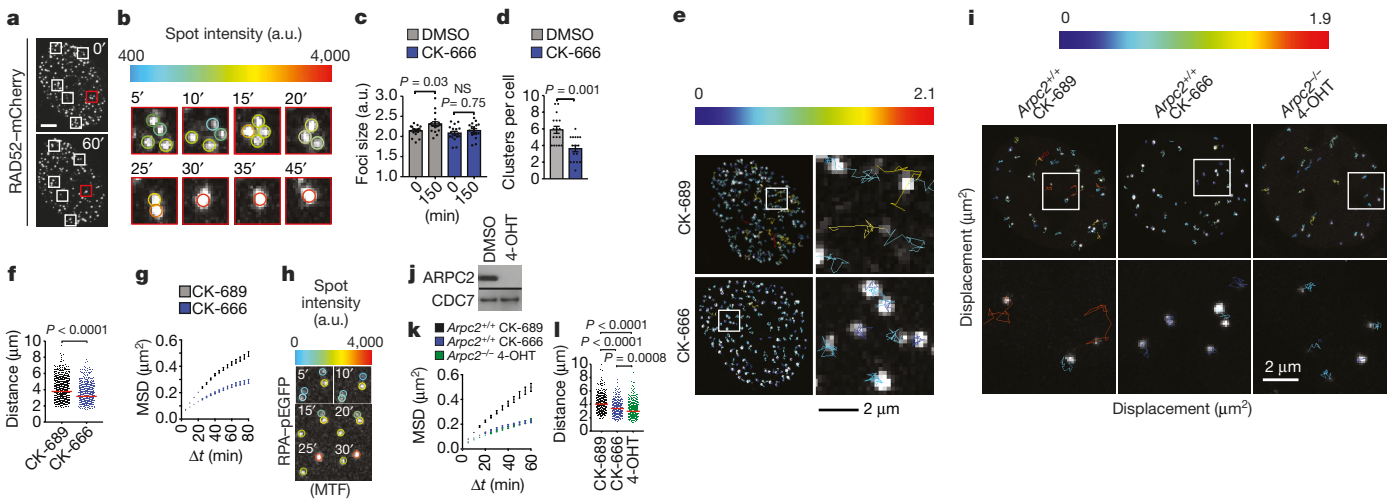


Fig. 3 | Arp2/3 drives DSB mobility during HDR. a, A representative U2OS cell nucleus with boxes indicating clustering of RAD52–mCherry foci. **b**, Expanded images of clustering from **a** (red box). Circles denote intensity of foci. **c**, Size of RAD52 foci (*P* calculated by one-way ANOVA with multiple comparisons; data shown as mean and s.e.m.; $n = 19$ nuclei (DMSO), 17 nuclei (CK-666). **d**, RAD52–mCherry clustering events (*P* calculated by two-tailed Mann–Whitney test; data shown as mean and s.e.m.; $n = 17$ nuclei (DMSO), 16 nuclei (CK-666). **e**, Traces of RAD52–mCherry foci over 100 min. Expanded images of boxed tracks (right). **f**, Median cumulative distance travelled by RAD52–mCherry foci. *P* calculated by two-tailed Mann–Whitney test; $n = 1{,}120$ foci from 13 nuclei (CK-689), 720 foci from 12 nuclei (CK-666). **g**, MSD of RAD52–mCherry foci. Data shown as mean and weighted s.e.m.; $n = 3{,}262$ foci from 13 nuclei (CK-689), 2,143 foci from 12 nuclei (CK-666). Δt, time interval. **h**, Clustering of pEGFP–RPA32–NLS foci in MTFs. **i**, Traces of RPA32–pEGFP–NLS foci over 100 min. **j**, ARPC2 protein levels in whole-cell lysates from Arpc2-LoxP-CreER MTFs (single experiment). **k**, MSD of RPA32–pEGFP–NLS foci. Data shown as mean and weighted s.e.m.; $n = 790$ foci from 13 nuclei (CK-689), 823 foci from 12 nuclei (CK-666), 1,135 foci from 13 nuclei (4-OHT). Scale bars, 5 μm or as indicated. **l**, Median cumulative distance travelled by RPA32–pEGFP–NLS foci. *P* calculated by one-way ANOVA with multiple comparisons; $n = 381$ foci from 13 nuclei (CK-689), 370 foci from 12 nuclei (CK-666), 425 from 13 nuclei (4-OHT).

in G1 cells (Fig. 2), the movement of 53BP1–YFP foci in G1 cells was unaffected by ARP2/3 inhibition (Extended Data Figs. 2a, 5a–c). In G2 cells, many 53BP1–YFP and RAD52 foci colocalized, consistent with the localization of 53BP1 at sites of DSB resection[22] (Extended Data Fig. 5d). The diffusion coefficient for 53BP1 foci increased from $2.4 \times 10^{-5}\,\mu m^2\,s^{-1}$ in G1 cells to $3.5 \times 10^{-5}\,\mu m^2\,s^{-1}$ in G2 cells (Extended Data Fig. 5e, f). Unlike in G1 cells, the motion of 53BP1 foci in G2 cells was reduced by ARP2/3 inactivation (Extended Data Fig. 5g). Collectively, these data indicate that ARP2/3 enhances the motion of sites of DSB resection in G2 cells.

ARP2/3 nucleates actin foci at sites of HDR

Genotoxic agents induce nuclear actin structures in mammalian cells[16]. In U2OS cells transfected with nuclear actin-chromobody TagGFP[25], we identified a subset of actin structures resembling foci (actin-cb foci); the number of these foci increased after NCS treatment (Fig. 4a, b and Extended Data Fig. 6a, b). We also identified brighter, rod-like filaments (Extended Data Fig. 6a, b). Strikingly, actin-cb foci clustered and merged over a 20-min interval (Fig. 4c and Supplementary Video 2a, b). MSD analysis of actin-cb foci revealed a $D(t)$ of $3.4 \times 10^{-4}\,\mu m^2\,s^{-1}$ (Fig. 4d, e). Similar to RAD52 and RPA, the α-coefficient for actin-cb

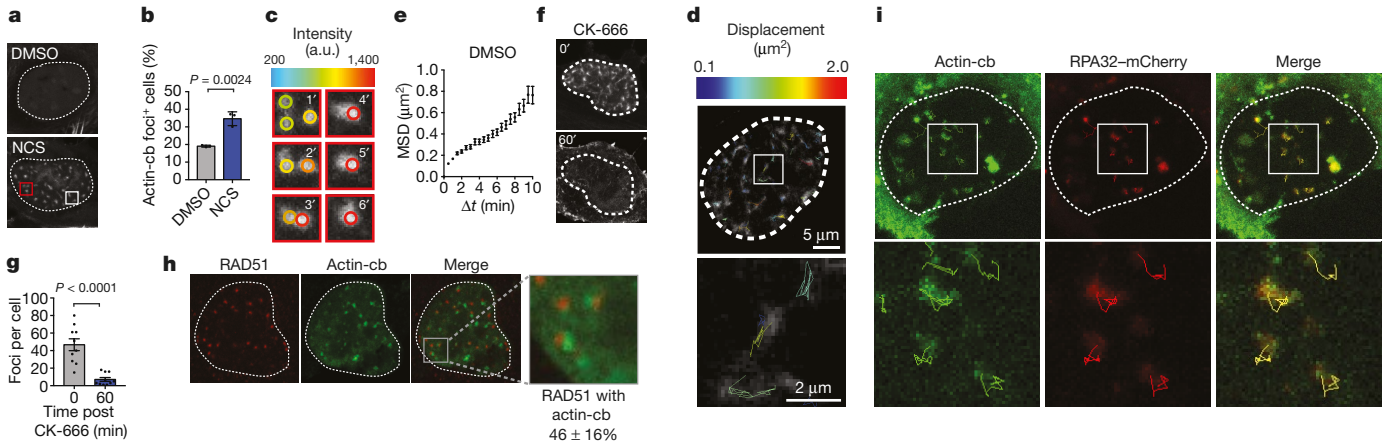


Fig. 4 | Nuclear actin foci cluster and localize to HDR sites. a, Representative images of U2OS nuclei transfected with nuclear actin-chromobody TagGFP. Boxes indicate clustering events. **b**, Percentage of cells with actin-cb foci from 3 biological replicates. *P* calculated by Student's two-tailed *t*-test; data shown as mean and s.d.; $n = 473$ nuclei (DMSO), 473 nuclei (NCS). **c**, Expanded images of a clustering event from **a** (red box). Circles denote intensity of foci. **d**, Representative traces of actin-cb foci (out of three independent experiments) over 10 min. **e**, MSD of actin-cb foci. Data shown as mean and weighted s.e.m.; DMSO $n = 662$ foci from 11 nuclei. **f**, Representative images of a U2OS nucleus with actin-cb foci following CK-666 treatment. **g**, Quantification of actin-cb foci before and after CK-666 treatment. *P* calculated by paired two-tailed *t*-test; data shown as mean and s.e.m.; $n = 10$ nuclei from two independent experiments. **h**, Representative images of a U2OS nucleus showing RAD51 co-localization with actin-cb foci. The percentage of RAD51 foci that overlap or touch actin-cb foci is shown ($n = 12$ nuclei). **i**, Representative images of a U2OS nucleus (from five independent experiments) showing traces of RPA32–mCherry–NLS foci (red) and actin-cb foci (green). Scale bars, 5 μm or as indicated.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

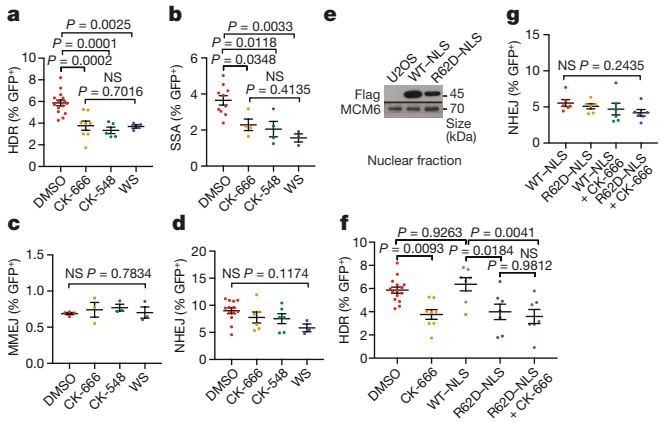


Fig. 5 | WASP and ARP2/3 mediate HDR. a, Summary of DR-GFP assay, which monitors recombination between two GFP repeats ($n = 15, 7, 5, 4$ left to right). NS, not significant. **b**, Summary of SA-GFP (SSA) assay ($n = 10, 4, 4, 3$ left to right). **c**, Summary of EJ2-GFP (MMEJ) assay ($n = 4, 3, 3, 3$, left to right). **d**, Summary of EJ5-GFP (NHEJ) assay ($n = 13, 6, 6, 3$, left to right). **e**, Western blot shows Flag-WT-NLS or Flag-R62D-NLS overexpression in U2OS nuclear soluble fraction (single experiment). **f**, DR-GFP assay in cells transfected with Flag-WT-NLS or Flag-R62D-NLS. ($n = 7, 7, 7$, left to right). HDR efficiency in the presence of DMSO or CK-666 (Fig. 5a) shown for comparison. **g**, EJ5-GFP assay in cells transfected with Flag-WT-NLS or Flag-R62D-NLS. ($n = 6, 6, 6$, left to right). For **a**–**g**, *P* calculated by one-way ANOVA with multiple comparisons as indicated; data shown as mean and s.e.m. *n* represents independent experiments. WS, Wiskostatin.

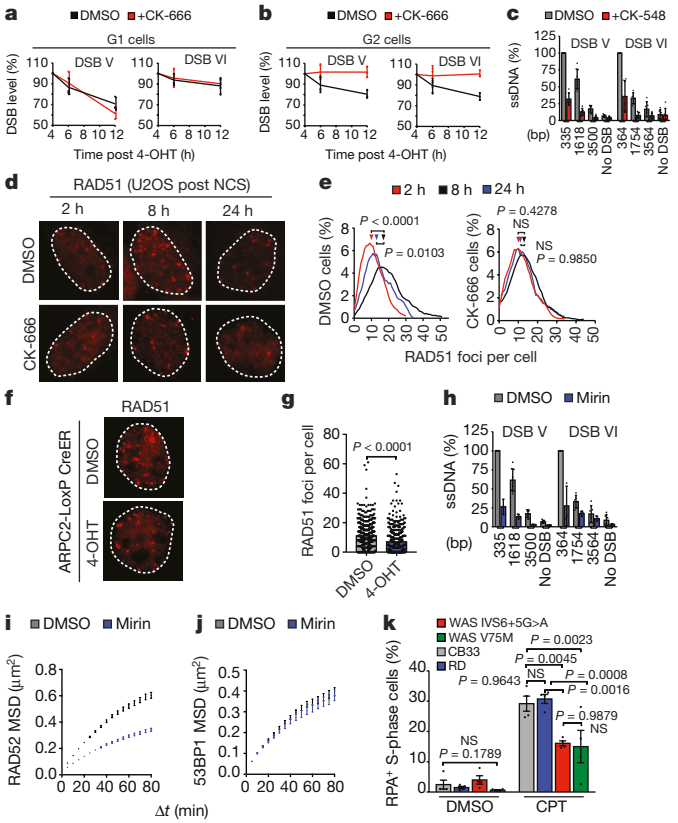


Fig. 6 | ARP2/3 facilitates resection and repair in G2. a, DSB repair at DSBs V and VI in G1-synchronized, ER-AsiSI-AID U2OS cells (tagged with an auxin-inducible degron (AID) that ablates the enzyme upon auxin treatment). Mean and s.d. ($n = 3$ technical replicates) of a representative experiment shown. **b**, DSB repair at DSBs V and VI in G2-synchronized ER-AsiSI-AID U2OS cells. Mean and s.d. ($n = 3$ technical replicates) of a representative experiment shown. **c**, Enrichment of ssDNA at DSBs V and VI following CK-548 treatment. Data shown as mean and s.d., $n = 6$ replicates from two independent experiments. **d**, Representative images of U2OS nuclei showing RAD51 foci. **e**, Quantification of RAD51 foci. *P* calculated by one-way ANOVA with multiple comparisons; smoothed traces show distribution of RAD51 foci. Arrows above curves indicate mean number of foci per cell. DMSO: 2 h $n = 1,830$ cells, 8 h $n = 2,440$ cells, 24 h $n = 782$ cells; CK-666: 2 h $n = 2,438$ cells, 8 h $n = 2,189$ cells, 24 h $n = 912$ cells. **f**, Representative images of ARPC2-LoxP-CreER MTF nuclei showing RAD51 foci. **g**, Quantification of RAD51 foci in MTFs. *P* calculated by two-tailed Mann–Whitney test; data shown as mean; $n = 571$ nuclei (DMSO), 462 nuclei (4-OHT). **h**, Representative ssDNA enrichment at DSBs V and VI following mirin treatment. Data shown as mean and s.d.; $n = 3$ replicates of two independent experiments. ssDNA in DMSO-treated cells (Fig. 5c) shown for comparison. **i**, MSD of RAD52–mCherry foci. Data shown as mean and weighted s.e.m.; DMSO $n = 3,292$ foci from 12 cells, mirin $n = 2,677$ foci from 11 cells. **j**, MSD of 53BP1–YFP foci in G1 cells. Data shown as mean and weighted s.e.m.; DMSO $n = 893$ foci from 12 cells, mirin $n = 744$ foci from 14 cells. **k**, Quantification of chromatin-bound RPA in S-phase cells. *P* calculated by two-way ANOVA with multiple comparisons; data shown as mean and s.e.m.; $n = 4$ independent experiments. WAS IVS6+5G>A and WAS V75M are cells carrying WAS mutations. RD and CB33, healthy lymphocytes. All DMSO columns, not significant ($P = 0.1789$). For CPT, CB33 versus RD, not significant ($P = 0.9643$); WAS IVS6+5G>A versus WAS V75M, not significant ($P = 0.9879$); CB33 versus WAS IVS6+5G>A, $P = 0.0045$; CB33 versus WAS V75M, $P = 0.0023$; RD versus WAS IVS6+5G>A, $P = 0.0016$; RD versus WAS V75M, $P = 0.0008$.

foci indicated confined motion. Notably, CK-666 abolished actin-cb foci, strongly suggesting that these structures are sites of ARP2/3-dependent nucleation (Fig. 4f, g, Supplementary Video 2c–e and Extended Data Fig. 6d). By contrast, nuclear actin rods exhibited limited mobility and were unaffected by CK-666 (Extended Data Fig. 6a–d). Finally, we compared the localization of actin-cb foci with RAD51 foci in fixed cells and RPA32–mCherry foci using live cell imaging. Following NCS treatment, 46% of RAD51 foci overlapped or abutted actin-cb foci (Fig. 4h). Moreover, we observed co-localization of actin-cb foci and RPA32–mCherry foci, which travelled together in the nucleoplasm (Fig. 4i and Extended Data Fig. 6e, f). Collectively, these experiments provide direct evidence that ARP2/3 assembles dynamic nuclear actin structures at sites of HDR.

WASP and ARP2/3 mediate HDR but not NHEJ

To determine how DSB clustering affects DNA repair pathways, we used a panel of U2OS cell lines to monitor the repair of DSBs induced by the bacterial endonuclease I-SceI by HDR, NHEJ, single-strand annealing (SSA), and microhomology-mediated end joining (MMEJ)[26]. Inactivation of ARP2/3 by CK-666 or CK-548 reduced the ability of cells to repair I-SceI-induced DSBs using HDR or SSA by 40% (Fig. 5a, b and Extended Data Fig. 7a, e). Similarly, wiskostatin, a small-molecule inhibitor of WASP[27], reduced the efficiency of HDR and SSA (Fig. 5a, b). Cells depleted of WASP by RNA interference showed similar defects in HDR (Extended Data Fig. 7b–d). Notably, inhibition of WASP or ARP2/3 did not compromise MMEJ or NHEJ, neither of which require substantial resection (Fig. 5c, d and Extended Data Fig. 7f, g). Finally, the application of CK-689 did not reduce the repair of I-SceI-induced DSBs by HDR, SSA or NHEJ (Extended Data Fig. 7h).

Although the ARP2/3 complex is found in the nucleus[14], it also polymerizes actin in the cytoplasm. To support the idea that the HDR defect was a direct consequence of inhibiting nuclear actin polymerization, we increased nuclear actin levels by overexpressing actin fused to a nuclear localization sequence (actin–NLS) (Fig. 5e). Whereas expression of wild-type actin–NLS did not affect HDR, overexpression of an actin variant that interrupts filament formation (actin(R62D)–NLS) inhibited HDR to levels comparable with ARP2/3 inhibition (Fig. 5f). Treatment of cells expressing actin(R62D)–NLS with CK-666 did not further reduce HDR, suggesting that nuclear actin and ARP2/3 functioned together. Moreover, expression of wild-type actin, actin(R62D), or wild-type actin–NLS tagged with mCherry did not affect DSB repair, whereas expression of actin(R62D)–NLS significantly inhibited HDR (Extended Data Fig. 8a, b). This strongly suggests that inhibition of

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

actin polymerization in the cytoplasm does not influence DNA repair in the nucleus. Similar to CK-666, expression of actin(R62D)–NLS did not reduce NHEJ-based repair (Fig. 5g). Finally, neither small interfering RNA (siRNA) directed against the actin filament nucleator formin-2 nor treatment with the formin inhibitor SMIFH2 inhibited HDR efficiency (Extended Data Fig. 8c–f). This observation demonstrates a specific requirement for ARP2/3 complex activity in HDR.

ARP2/3 activity enhances DSB resection

To assess the functional importance of ARP2/3-driven DSB mobility, we compared the repair kinetics of two AsiSI-induced DSB sites (DSBs V and VI) that undergo up to 3.5 kb of resection[28]. In G1 cells, the repair kinetics of both DSBs were identical in the presence or absence of CK-666, confirming that ARP2/3 does not influence NHEJ (Fig. 6a). By contrast, in G2 cells, CK-666 substantially slowed repair at these DSB sites (Fig. 6b).

To investigate how actin-driven clustering might promote HDR efficiency, we tested whether the ARP2/3 complex facilitates resection. Inactivation of the ARP2/3 complex decreased ssDNA generation at DSBs V and VI by 50%, indicating that clustering facilitates DSB end-resection (Fig. 6c). We consolidated this finding by quantifying RAD51 foci that arose following DSB resection in NCS-treated cells over 24 h (Fig. 6d, e). In control cells, the number of RAD51 foci first increased between 2 and 8 h and then decreased up to 24 h, indicative of recruitment and repair (Fig. 6d). By contrast, RAD51 foci increased minimally following ARP2/3 inhibition (Fig. 6d, e). Moreover, substantially fewer RAD51 foci arose in ARPC2-depleted MTFs than in control MTFs (Fig. 6f, g).

In yeast and human cells, DSB motion requires the HDR machinery[1,2,5]. We wondered whether RAD52 movement requires the activity of the MRN (MRE11, RAD50 and NBS1) complex[29]. Accordingly, we treated U2OS cells with mirin, a small-molecule inhibitor that abolishes the nuclease activity of MRE11[30]. Mirin lowered ssDNA levels at DSBs V and VI by 70% (Fig. 6h). Notably, mirin substantially decreased the $D(t)$ of RAD52 foci from $4.4 \times 10^{-5}\,\mu m^2\,s^{-1}$ to $2.4 \times 10^{-5}\,\mu m^2\,s^{-1}$ (Fig. 6i and Extended Data Fig. 2a), but did not significantly affect the MSD of 53BP1 foci in G1 cells (Fig. 6j). Together, these observations suggest a positive feedback loop, wherein resection enhances DSB movements, which, in turn, enhance resection.

WAS is an X-linked disorder characterized by severe immunodeficiency and predisposition to non-Hodgkin's lymphoma and leukaemia[31]. Given that inactivation of WASP by wiskostatin or siRNA produced defects in HDR (Fig. 5a and Extended Data Fig. 7d), we tested B lymphocytes derived from patients with WAS for defects in DSB end-resection. Relative to lymphocytes derived from healthy individuals, lymphocytes bearing two distinct WAS mutations exhibited 50% fewer RPA-positive cells upon exposure to the DNA topoisomerase I inhibitor camptothecin (CPT), consistent with a resection defect (Fig. 6k and Extended Data Fig. 9a–d). Moreover, WAS mutant cells were more sensitive to CPT treatment than healthy lymphocytes (Extended Data Fig. 10a, b). Inhibition of ARP2/3 by CK-666 also enhanced CPT-induced lethality in U2OS cells (Extended Data Fig. 10c). Finally, inhibition of ARP2/3 sensitized cells to the DNA polymerase inhibitor aphidicolin and the PARP1 inhibitor olaparib, indicating a reduced tolerance to replication stress (Extended Data Fig. 10d, e). Together, these results demonstrate that activation of ARP2/3 by WASP enhances the processing and resolution of DSBs undergoing HDR.

Discussion

DSB mobility is influenced by cell cycle stage, chromatin state, and repair pathway choice. Our finding that ARP2/3 specifically enhances the movement of breaks undergoing HDR is supported by studies showing that DSB mobility requires DNA end-resection. DSBs induced in pericentric heterochromatin during the S and G2 phases of the cell cycle relocate to the nuclear periphery, whereas DSBs generated during the G1 phase remain stable[32]. Critically, blocking HDR or resection prevented DSB mobility in G2. Similarly, in *Drosophila*, the activities of CtIP, Exo1, and Blm were necessary for DSB movement outside the heterochromatic domain[33]. Notably, DSBs within heterochromatin in *Drosophila* relocalize to repair sites along nuclear actin filaments in an ARP2/3-dependent process[34]. Conversely, we now show that inhibition of ARP2/3 decreases end- processing, consistent with positive feedback regulation between DSB mobility and resection. Indeed, DSB clustering could facilitate enzymatic reactions by increasing the local concentration of repair factors. Notably, in budding yeast, Arp2, Arp3, and Arpc2 are synthetic lethal in combination with Mre11 or Sgs1 deficiency[35]. This supports interdependency of actin polymerization and resection.

Our ChIP analyses suggest a two-step mechanism for actin nucleation at HDR sites (Extended Data Fig. 10f). WASP is recruited at all DSBs regardless of cell cycle stage, reminiscent of NHEJ factors. By contrast, ARP2/3 is recruited in G2 to breaks that assemble HDR machinery. Actin polymerization promoted by WASP and ARP2/3 might therefore require sequential activation of PI(3)K-like kinases. Whereas ATM promotes the assembly of complexes required for resection, ATR commits repair to HDR[29]. Accordingly, following ATM activation and WASP recruitment at DSBs, resection initiation could create a permissive environment for ARP2/3 activity leading to increased chromosomal mobility at specific DSB sites.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0237-5.

Received: 5 October 2017; Accepted: 17 April 2018;
Published online 20 June 2018.

1. Dion, V., Kalck, V., Horigome, C., Towbin, B. D. & Gasser, S. M. Increased mobility of double-strand breaks requires Mec1, Rad9 and the homologous recombination machinery. *Nat. Cell Biol.* **14**, 502–509 (2012).
2. Miné-Hattab, J. & Rothstein, R. Increased chromosome mobility facilitates homology search during recombination. *Nat. Cell Biol.* **14**, 510–517 (2012).
3. Lisby, M., Mortensen, U. H. & Rothstein, R. Colocalization of multiple DNA double-strand breaks at a single Rad52 repair centre. *Nat. Cell Biol.* **5**, 572–577 (2003).
4. Aten, J. A. et al. Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains. *Science* **303**, 92–95 (2004).
5. Aymard, F. et al. Genome-wide mapping of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes. *Nat. Struct. Mol. Biol.* **24**, 353–361 (2017).
6. Lemaître, C. et al. Nuclear position dictates DNA repair pathway choice. *Genes Dev.* **28**, 2450–2463 (2014).
7. Robinett, C. C. et al. *In vivo* localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J. Cell Biol.* **135**, 1685–1700 (1996).
8. Soutoglou, E. et al. Positional stability of single double-strand breaks in mammalian cells. *Nat. Cell Biol.* **9**, 675–682 (2007).
9. Cho, N. W., Dilley, R. L., Lampson, M. A. & Greenberg, R. A. Interchromosomal homology searches drive directional ALT telomere movement and synapsis. *Cell* **159**, 108–121 (2014).
10. Lottersberger, F., Karssemeijer, R. A., Dimitrova, N. & de Lange, T. 53BP1 and the LINC complex promote microtubule-dependent DSB mobility and DNA repair. *Cell* **163**, 880–893 (2015).
11. Virtanen, J. A. & Vartiainen, M. K. Diverse functions for different forms of nuclear actin. *Curr. Opin. Cell Biol.* **46**, 33–38 (2017).
12. Taylor, M. D. et al. Nuclear role of WASp in the pathogenesis of dysregulated TH1 immunity in human Wiskott–Aldrich syndrome. *Sci. Transl. Med.* **2**, 37ra44 (2010).
13. Wu, X. et al. Regulation of RNA-polymerase-II-dependent transcription by N-WASP and its nuclear-binding partners. *Nat. Cell Biol.* **8**, 756–763 (2006).
14. Yoo, Y., Wu, X. & Guan, J. L. A novel role of the actin-nucleating Arp2/3 complex in the regulation of RNA polymerase II-dependent transcription. *J. Biol. Chem.* **282**, 7616–7623 (2007).
15. Pollard, T. D. & Borisy, G. G. Cellular motility driven by assembly and disassembly of actin filaments. *Cell* **112**, 453–465 (2003).
16. Belin, B. J., Lee, T. & Mullins, R. D. DNA damage induces nuclear actin filament assembly by Formin-2 and Spire-1/2 that promotes efficient DNA repair. *eLife* **4**, e07735 (2015).
17. Hetrick, B., Han, M. S., Helgeson, L. A. & Nolen, B. J. Small molecules CK-666 and CK-869 inhibit actin-related protein 2/3 complex by blocking an activating conformational change. *Chem. Biol.* **20**, 701–712 (2013).
18. Nolen, B. J. et al. Characterization of two classes of small molecule inhibitors of Arp2/3 complex. *Nature* **460**, 1031–1034 (2009).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

19. Rogakou, E. P., Boon, C., Redon, C. & Bonner, W. M. Megabase chromatin domains involved in DNA double-strand breaks *in vivo*. *J. Cell Biol*. **146**, 905–916 (1999).
20. Aymard, F. et al. Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat. Struct. Mol. Biol*. **21**, 366–374 (2014).
21. Karanam, K., Kafri, R., Loewer, A. & Lahav, G. Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase. *Mol. Cell* **47**, 320–329 (2012).
22. Ochs, F. et al. 53BP1 fosters fidelity of homology-directed DNA repair. *Nat. Struct. Mol. Biol*. **23**, 714–721 (2016).
23. Rotty, J. D. et al. Profilin-1 serves as a gatekeeper for actin assembly by Arp2/3-dependent and -independent pathways. *Dev. Cell* **32**, 54–67 (2015).
24. Caron, P. et al. Non-redundant functions of ATM and DNA-PKcs in response to DNA double-strand breaks. *Cell Rep.* **13**, 1598–1609 (2015).
25. Plessner, M., Melak, M., Chinchilla, P., Baarlink, C. & Grosse, R. Nuclear F-actin formation and reorganization upon cell spreading. *J. Biol. Chem*. **290**, 11209–11216 (2015).
26. Gunn, A. & Stark, J. M. I-SceI-based assays to examine distinct repair outcomes of mammalian chromosomal double strand breaks. *Methods Mol. Biol*. **920**, 379–391 (2012).
27. Peterson, J. R. et al. Chemical inhibition of N-WASP by stabilization of a native autoinhibited conformation. *Nat. Struct. Mol. Biol*. **11**, 747–755 (2004).
28. Zhou, Y., Caron, P., Legube, G. & Paull, T. T. Quantitation of DNA double-strand break resection intermediates in human cells. *Nucleic Acids Res*. **42**, e19, doi:10.1093/nar/gkt1309 (2014).
29. Symington, L. S. & Gautier, J. Double-strand break end resection and repair pathway choice. *Annu. Rev. Genet*. **45**, 247–271 (2011).
30. Dupré, A. et al. A forward chemical genetic screen reveals an inhibitor of the Mre11-Rad50-Nbs1 complex. *Nat. Chem. Biol*. **4**, 119–125 (2008).
31. Massaad, M. J., Ramesh, N. & Geha, R. S. Wiskott-Aldrich syndrome: a comprehensive review. *Ann. NY Acad. Sci*. **1285**, 26–43 (2013).
32. Tsouroula, K. et al. Temporal and spatial uncoupling of DNA double strand break repair pathways within mammalian heterochromatin. *Mol. Cell* **63**, 293–305 (2016).
33. Chiolo, I. et al. Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. *Cell* **144**, 732–744 (2011).
34. Caridi, C. P. et al. Nuclear F-actin and myosins drive relocalization of heterochromatic breaks. *Nature* https://doi.org/10.1038/s41586-018-0242-8 (2018).
35. van Pel, D. M., Stirling, P. C., Minaker, S. W., Sipahimalani, P. & Hieter, P. *Saccharomyces cerevisiae* genetics predicts candidate therapeutic genetic interactions at the mammalian replication fork. *G3* **3**, 273–282 (2013).

Acknowledgements We thank T. Swayne and E. L. Munteanu from the Confocal and Specialized Microscopy Shared Resource of the Irving Cancer Research Center (ICRC) and the Flow Cytometry Shared Resource of the ICRC at Columbia University (supported by NIH grants #P30-CA013696, #S10-RR025686); J. Stark for the GFP reporter cell lines; G. Legube for the ER-AsiSI U2OS lines and for sharing unpublished data; J. Bear for the ARPC2-LOXP-CreER MTFs; G. Lahav for the RAD52–mCherry U2OS line; R. Dalla-Favera for the RD and CB33 cell lines; J. Lukas for the pEGFP–NLS–RPA32 construct; R. Baer for the I-SceI plasmid; N. Kato and T. Lagache for manuscript comments; and A. Hollar for Metacyte classifier design. This work was supported by the NIH (R35-CA197606 and P01-CA174653 to J.G., RO1-GM099481 to G.G.G., PHS-GM103314 and PHS-GM109824 to B.T.C., and F30-CA217049 to B.R.S.).

Author contributions J.G. and B.R.S conceived the study and wrote the manuscript. B.R.S. conducted the majority of the experiments and data analyses. T.A. and Y.L. performed the mass spectrometry experiments. G.G.G., M.E.G., W.C., B.T.C., T.A. and Y.L. assisted with data analysis and interpretation.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0237-5.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0237-5.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to J.G.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Chromatin binding assay and western blot. Preparation of low-speed supernatant (LSS) *Xenopus* egg extracts and isolation of demembranated sperm nuclei (chromatin) were performed as previously described[36]. For chromatin binding experiments, 15 μl LSS extract was supplemented with demembranated sperm nuclei (2,500 sperm per μl) and incubated for 10 min at 21 °C before addition of DMSO or one of the following compounds for 10 min: CK-666 (160 μM), CK-548 (160 μM), KU55933 (100 μM), VE821 (50 μM). Subsequently, the PflMI restriction endonuclease was added to each sample (0.05 U/μl) before incubation for an additional 60 min at 21 °C. Extracts were diluted in chromatin isolation buffer (50 mM HEPES-KOH, pH 7.8, 100 mM KCl, 2.5 mM $MgCl_2$) supplemented with 0.125% Triton X-100 and overlaid on top of sucrose cushions (chromatin isolation buffer plus 30% sucrose) in 1.5-ml low-retention tubes (Thermo Fisher Scientific). Samples were spun at 7,180 r.p.m. for 30 min at 4 °C in a swing-bucket Sorval rotor (HB-6). Chromatin pellets were resuspended in 10 μl Laemmli buffer, boiled, and fractionated on a 4–22% Tris-glycine gel (Invitrogen) according to standard procedures, followed by transfer of resolved proteins onto PVDF membranes. Following a 1-h block with 5% milk, membranes were incubated overnight at 4 °C with one of the following primary antibodies: β-actin (Sigma Aldrich: A5316, 1/1,000), ARPC4 (Novus Biologicals: NBP1-69003, 1/500), capping protein β-2 subunit (Developmental Studies Hybridoma Bank: E00007, 1/10,000), histone H3 (Cell Signaling Technology, 9715, 1/2,000) and *Xenopus* Mre11[36]. HRP-conjugated secondary antibodies were used (anti-rabbit IgG HRP, anti-mouse IgG HRP, Fisher Scientific) and chemiluminescence (Supersignal West Pico Chemiluminescent Substrate, 34077) was used.

Mass spectrometric analysis of DSB-containing chromatin. Chromatin samples were isolated as described above with the following modifications: 100 μl LSS extract was supplemented with sperm nuclei (5,000 sperm per μl) and incubated for 10 min at 21 °C before addition of PflMI restriction endonuclease (0.05 U/μl) and incubated for an additional 60 min. Chromatin was isolated through sucrose cushions as described above and extensively digested with 100 U micrococcal nuclease for 30 min at 37 °C (NEB). Chromatin proteins were reduced with dithiothreitol (5 mM, 30 min 60 °C) and alkylated with iodoacetamide (15 mM, 30 min, room temperature) before fractionation on SDS–PAGE. 'In-gel' digestion was performed with proteomic-grade trypsin (Promega) at 5 ng/μl for 16 h at 37 °C. The resulting digestion peptides were extracted from the gel pieces with 1:2 (vol:vol) 5% formic acid: acetonitrile solution. The resulting peptides were labelled with isobaric mass tags (ITRAQ 4-plex, Sciex), combined, purified with in-house made STAGE tips[37], resuspended in 0.5% acetic acid and loaded onto a home-packed reverse phase C18 column (75 μm ID). The peptides were separated using a linear gradient (0–42% acetonitrile, 0.5% acetic acid, 120 min, 150 nl/min) and directly sprayed into an LTQ-Orbitrap-Velos mass spectrometer for analysis (Thermo). The repetitive analytical cycle incorporated a high-resolution mass scan in the Orbitrap (resolution 30,000) followed by tandem MS scans of the five most intense peaks observed in each Orbitrap mass spectrum. Peptides were identified and quantified using Proteome Discoverer software (Thermo, Version 1.4).

Cell culture and drug treatment. U2OS and mouse-tail fibroblast cell lines were cultured in high-glucose Dulbecco's modified Eagle's medium supplemented with L-glutamine, 10% fetal bovine serum, and 1% penicillin-streptomycin. For neocarzinostatin-dependent DSB induction, cells were cultured on 8-well chamber slides (Thermo Fisher Scientific) and were treated with 0.5 μg/ml NCS for 60 min at 37 °C. Cells were washed twice with PBS before the addition of medium containing DMSO or one of the following compounds: CK-666 (Sigma Aldrich: SML-006, 100 μM), CK-548 (Sigma Aldrich: C7499, 50 μM), CK-689 (Sigma Aldrich: 18251750, 100 μM), and wiskostatin (Sigma Aldrich: W2270, 3 μM). After drug treatment, cells were then incubated at 37 °C for the indicated times. For mirin experiments, cells were pre-treated with mirin (Sigma Aldrich: M9948, 50 μM) for 1 h, before addition of NCS.

ER-AsiSI U2OS cells stably expressing the AsiSI restriction endonuclease tethered to the oestrogen receptor were provided by G. Legube (Centre de Biologie Integrative, Toulouse). For cell synchronization, cells were treated with 2 mM thymidine for two 18-h intervals separated by an 11-h release in fresh medium. Cells underwent double-thymidine block and were released into fresh medium for 7 h (G2) or 15 h (G1) before AsiSI-dependent DSB induction with medium supplemented with 300 nM 4-OHT (Sigma Aldrich, H7904) plus CK-666, CK-548, wiskostatin, or DMSO. Cells were subsequently incubated at 37 °C for 4 h.

Arpc2$^{-/-}$ fibroblast cell lines (provided by J. E. Bear, University of North Carolina) stably express the Cre-recombinase tagged to the oestrogen receptor. Upon 4-OHT treatment, the *Arpc2* locus is floxed by Cre and cells are depleted of ARPC2 following 4 days of 4-OHT treatment.

CB33 and RD lymphoblastoid B cell lines were a gift from R. Dalla-Favera (Columbia University). Two B lymphocyte cell lines bearing distinct mutations in the *WAS* gene were obtained from the NIGMS Human Genetic Cell Repository (Coriell: GM1267, GM1268). GM1267 harbours a G>A transition at position 4 in intron 6 of the *WAS* gene (IVS6+5G>A). This mutation activates a cryptic splice site leading to splicing of a 38-nucleotide sequence from intron 6 onto exon 7. GM1268 harbours a G>A transition at nucleotide 257 in exon 2 on the WAS gene. This mutation results in a methionine-for-valine substitution at codon 75 (V75M). All B cell lines were grown in Rosewell Park Memorial Institute medium with 20% heat-inactivated bovine serum with 2 mM L-glutamine added. All cell lines were tested for mycoplasma contamination.

Immunohistochemistry and quantification of DSB repair protein foci. U2OS cells were cultured on 8-well chamber slides (Thermo Fisher Scientific) and subjected to drug treatments as described above. Cells were then washed once with PBS before fixation with freshly prepared 4% PFA (pH 7.4) for 10 min. Cells were subsequently washed twice with PBS for 10 min before permeabilization with 0.1% PBS-Triton X-100 for 10 min. Cells were then washed once with PBS for 5 min before incubation with blocking buffer (3% BSA in 0.2% PBS-Tween) for 1 h. Cells were stained with primary antibodies diluted in blocking buffer under a Hybrislip (Invitrogen) overnight in a humidified chamber at 4 °C. Primary antibodies include γH2AX (EMD Millipore: 05-636, 1/500 or Abcam: ab81299, 1/500), WASP (Santa Cruz: sc-5300, 1/50), RAD51 (Santa Cruz: sc-8349, 1/50), 53BP1 (Novus Biologicals: NC100-304, 1/250) and Flag (Sigma Aldrich: F3165, 1/250). Cells were then washed three times in PBS for 15 min and incubated with a fluorescence-conjugated secondary antibody in PBS with DAPI (Invitrogen, 1/10,000) for 1 h at room temperature. Secondary antibodies were Alexa 488 conjugated goat anti-mouse Ig (Abcam: ab150113, 1/1,000), Alexa 488 conjugated goat anti-rabbit Ig (Thermo Fisher Scientific, A-11034, 1/10,000), and Alexa 594 conjugated goat anti-mouse Ig (Thermo Fisher Scientific, A-11005, 1/10,000). Cells were then washed three times with PBS for 15 min. Vectashield was applied to each slide and slides were coverslipped. Slides were analysed under 40× magnification using a Zeiss Axio Imager Z2 microscope, equipped with a CoolCube1 camera (Carl Zeiss). Images were processed for contrast enhancement and background reduction using ImageJ.

MetaCyte software (Metasystems, version 3.10.6) was used to detect U2OS nuclei based on DAPI staining and to perform automated quantification of γH2AX, RAD51, and WASP foci within each nucleus. For analysis of the size and clustering of foci, images were converted to greyscale and processed with Icy Software (Version 1.8.6.0; Institut Pasteur, Quantitative Image Analysis Unit; http://www.icy.bioimageanalysis.org). Regions of interest were drawn around individual nuclei and foci were detected using the Spot Detector plug-in (Channel 0)[38]. For clustering analyses, spots detected using Scale 2 (3 pixels) were exported to the spatial analysis plug-in, which uses Ripley's *K* function to assess the deviance of pairs of points from total randomness[24,38]. The software reports statistically significant clustering ($P < 0.05$) when the *K* function crosses the clustering threshold. The software further corrects for the number of points in the computations of the quantiles of the Ripley's *K* function. Thus, there is no loss of statistical power for samples with lower numbers of foci. For foci size analysis, foci with minimum sizes of 1 and 3 pixels were counted using the Spot Detector plug-in. Scale 1 was used to detect foci that were at least 1 pixel in size, whereas Scale 2 was used to detect foci that were at least 3 pixels. Foci that met criteria for both Scale 1 and Scale 2 were excluded from the Scale 1 group to avoid duplicate detections. Foci counts were subsequently averaged to determine the mean size of foci per nucleus (100 nuclei were detected per condition or as indicated). For co-localization studies, images were imported into Image J (NIH) and separated into green (γH2AX) and red (WASP) channels. These files were subsequently imported into Icy and regions of interest were drawn around individual nuclei. Pearson's *r* was calculated using the Co-localization Studio plug-in, which compares the coincidence of spots inside a region of interest between green and red channels.

Chromatin immunoprecipitation. ChIP experiments followed previously published protocols[20] with the following modifications: 50 million ER-AsiSI U2OS cells were cultured on 150-mm^2 plates in high-glucose Dulbecco's modified Eagle's medium supplemented with L-glutamine, 10% fetal bovine serum, and 1% penicillin-streptomycin. Following double-thymidine block, cells were released into fresh medium for 7 h (G2) or 15 h (G1) before incubation with 300 nM 4-OHT for 4 h. Cells were then trypsinized, washed with PBS, cross-linked in 1% methanol-free formaldehyde fixing buffer for 10 min, quenched with 0.125 M glycine for 5 min, and snap frozen at −80 °C. Cell lysis and nuclear isolation were performed using NP40 lysis buffer and SDS shearing buffer (Covaris). Nuclei were sonicated using the S220 Ultrasonicator (Covaris) to obtain chromatin fragments of 500–1,000 bp in length. Sheared chromatin was incubated with 10 μg antibodies to RAD51 (Santa Cruz: sc-8349), DNA-PKcs (Abcam: ab1832), WASP (Santa Cruz: sc-5300), ARPC2 (Santa Cruz: sc-32195) or IgG (Jackson ImmunoResearch Laboratories) overnight. These antibodies were validated for ChIP by previous studies[12,20]. Protein A/G magnetic beads were added overnight, followed by sequential washes at increasing stringency and reverse cross-linking at 65 °C. Immunoprecipitated DNA and input DNA were analysed in triplicate

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

by qPCR. ChIP efficiencies (measured as per cent of input immunoprecipitated) were measured by qPCR at sites 80 bp downstream of DSBs. Graphs show ChIP efficiencies in G1, G2, or no 4-OHT cells, pooled from three experiments, and normalized against input material. Sequences for primers are listed below: DSB I: 5′-GTCCCTCGAAGGGAGCAC-3′, 5′-CCGACTTTGCTGTGTGACC-3′; DSB II: 5′-CCGCCAGAAAGTTTCCTAGA-3′, 5′-CTCACCCTTGCAGCACTTG-3′; DSB III: 5′-TCCCCTGTTTCTCAGCACTT-3′, 5′- CTTCTGCTGTTCTGCGTCCT-3′; DSB IV: 5′-ATCGGGCCAATCTCAGAGG-3′, 5′-GCGACGCTAACGTTA AAGCA-3′. Primer pairs used to measure RAD51, DNA-PKcs, WASP, and ARPC2 were located 80 bp from DSB sites or at an undamaged genomic locus.

I-SceI-induced DSB repair assays. U2OS cell lines which harbour chromosomally integrated I-SceI-based GFP reporter substrates were used to monitor the efficiency of HDR, SSA, MMEJ, and c-NHEJ repair processes. The DR-GFP, SA-GFP, EJ2-GFP, and EJ5-GFP cell lines (gifts from J. Stark, City of Hope Cancer Center) were transfected with 1 μg I-SceI-expressing vector (FuGENE 6 transfection reagent (Promega), Opti-MEM (Gibco)) in medium supplemented with DMSO or one of the following compounds: CK-666 (Sigma Aldrich: SML-006, 50 μM), CK-548 (Sigma Aldrich: C7499, 25 μM), CK-689 (Sigma Aldrich: 182517, 50 μM), SMIFH2 (Sigma Aldrich: S4826, 5 μM), or wiskostatin (Sigma Aldrich: W2270, 3 μM). The I-SceI-expressing vector was a gift from R. Baer. For the Flag actin construct overexpression experiments, DR-GFP cells were transfected with Flag–actin–NLS or Flag–actin(R62D)–NLS for 48 h before transfection with I-SceI in medium supplemented with DMSO or CK-666. For the mCherry actin construct overexpression experiments, DR-GFP cells were co-transfected with I-SceI and mCherry–actin, mCherry–actin–NLS, mCherry–actin(R62D), or mCherry– actin(R62D)–NLS. Cells were harvested 48 h after I-SceI transfection and percentage GFP^+ events were counted by fluorescence-activated cell sorting (FACS) (10,000 cells per biological replicate). For all conditions, parallel transfection with 1 μg pEGFP-N3 vector (Clontech) was used to determine transfection efficiency. Percentage refers to the number of GFP^+ cells divided by the number of $pEGFP^+$ cells.

Resection assay. A detailed protocol for the resection assay has been recently described[28]. In brief, ER-AsiSI U2OS cells were treated with 300 nM 4-OHT in medium supplemented with DMSO, CK-548 (50 μM), or CK-666 (100 μM). For mirin experiments, cells were pre-treated with mirin (Sigma Aldrich: M9948, 50 μM) for 1 h before addition of 4-OHT, and subsequently washed and incubated with mirin for the indicated times. After 4 h, cells were trypsinized and washed with cold PBS before resuspension in lysis buffer (100 mM NaCl, 10 mM TrisCl pH 8, 25 mM EDTA, pH 8, 0.5% SDS, 0.1 mg/ml proteinase K). Cells were agitated at 37 °C overnight. After phenol/chloroform purification and salt/ethanol extraction, precipitated DNA underwent restriction enzyme digestion by BamHI-HF, BsrGI, or HindIII-HF (1 unit enzyme per 7 ng DNA) at 37 °C overnight. DNA digests were subsequently used as templates in a 20 μl qPCR reaction containing 10.0 μl 2× TaqMan Universal PCR Master mix (Thermo), 0.5 μM of each primer, and 0.2 μM of probe. The extent of ssDNA generation at sites downstream of various DSBs was determined by calculating the ΔCt value for each site (mock-digested Ct value − digest Ct value) and using the following equation: $\text{ssDNA} = 1/(2^{(\Delta Ct - 1) + 0.5)} \times 100$. Sequences for primers and probes are listed below: DSB V (335 bp site): 5′-GAATCGGATGTATGCGACTGATC-3′, 5′-TTCCAAAGTTATTCCAACCCGAT-3′; 6FAM-CACAGCTTGCCCATCCT TGCAAACC-TAMRA; DSB V (1618 bp site): 5′- TGAGGAGGTGACATTAGAA CTCAGA-3′, 5′-AGGACTCACTTACACGGCCTTT-3′; 6FAM-TTGCAAGG CTGCTTCCTTACCATTCAA-TAMRA; DSB V (3500 bp site): 5′-TCCTAGCCAG ATAATAATAGCTATACAAACA-3′, 5′-TGAATAGACAGACAACAGATAAA TGAGACA-3′; 6FAM-ACCCTGATCAGCCTTTCCATGGGTTAAG-TAMRA; DSB VI (364 bp site): 5′-CCAGCAGTAAAGGGGAGACAGA-3′, 5′-CTGTTCAATCGTCTGCCCTTC-3′; 6FAM-CCAGGCCCTCAAAATC CCTCCACTG-TAMRA; DSB VI (1754 bp site): 5′- GAAGCCATCCTACTCTTC TCACCT-3′, 5′-GCTGGAGATGATGAAGCCCA-3′; 6FAM-CACTCCCTGTTC TTCTTCTGCTCCCGA-TAMRA; DSB VI (3564 bp site): 5′-GCCC AGCTAAGATCTTCCTTCA-3′, 5′-CTCCTTTGCCCTGAGAAGTGA-3′; 6FAM-CTGCAGCCCTCAAGCCCGGAT-TAMRA. No DSB site: 5′- ATTG GGTATCTGCGTCTAGTGAGG-3′, 5′-GACTCAATTACATCCCTGCAGCT-3′; 6FAM-TCTCTGCACAGACCGGCTTCCCTTC-TAMRA.

AsiSI-AID repair assay. The repair assay followed recently published protocols with the following modifications[20,28]. Five million ER-AsiSI-AID U2OS cells were cultured on 100-mm^2 plates in high-glucose Dulbecco's modified Eagle's medium supplemented with L-glutamine, 10% fetal bovine serum, and 1% penicillin-streptomycin. Following double-thymidine block, cells were released into fresh medium for 7 h (G2) or 15 h (G1) before incubation with 300 nM 4-OHT and DMSO or CK-666 (100 μM) for 4 h. Cells were subsequently washed three times with PBS and then incubated in medium containing auxin (Sigma: I5148, 500 μg/ml) and DMSO or CK-666. Cells were harvested 0, 2, or 8 h after auxin exposure. Phenol/chloroform purification followed by salt/ethanol extraction was performed to precipitate DNA. DNA were subsequently used as templates in a 20-μl qPCR reaction containing 10.0 μl 2× TaqMan Universal PCR Master mix (Thermo), 0.5 μM of each primer, and 0.2 μM of probe. The level of unrepaired DSBs was determined by normalizing the Ct value for each DSB site with a control, undamaged site. Sequences for primers and probes are listed below. DSB V (across DSB): 5′-GATGTGGCCAGGGATTGG-3′, 5′-CACTCAAGCCCAACCCGT-3′; DSB VI (across DSB): 5′- GAGGAGCCTCTCCTGCAGC-3′, 5′-GAACCAGACCTAC CTCCAGGG-3′.

Live cell imaging. U2OS cells stably expressing RAD52–mCherry, 53BP1–YFP, and Geminin–CFP constructs[21] were cultured on 35-mm glass bottom microwell dishes (MatTek, P35GC-1.5-10-C). Cells were treated with 0.5 μg/ml NCS to induce DSBs for 60 min at 37 °C. Subsequently, cells were washed twice with PBS before the addition of medium containing DMSO, CK-689 (100 μM), or CK-666 (100 μM). For mirin experiments, cells were pre-treated with mirin (Sigma Aldrich: M9948, 50 μM) for 1 h before addition of NCS, and subsequently washed and incubated with mirin for the indicated times. S-phase cells were selected for RAD52 foci analysis by screening for Geminin–CFP positivity. G1 cells were selected for 53BP1 foci analysis by screening for Geminin–CFP^-/RAD52–$mCherry^-$ cells. After 20 h, images were acquired on an A1RMP confocal microscope (Nikon Instruments), on a TiE Eclipse stand equipped with a 60×/1.49 Apo-TIRF oil-immersion objective lens, an automated XY stage, stage-mounted piezoelectric focus drive, and a heated, humidified stagetop chamber with 5% CO_2 atmosphere. Cells expressing RAD52–mCherry and 53BP1–YFP foci were imaged in GaAsP detectors using 561 nm and 488 nm excitation, respectively, and standard RFP and GFP emission filters. The confocal pinhole was set to 1 Airy unit for the red channel. Foci movements were examined by collecting *z* series at 0.4-μm intervals throughout the entire nucleus every 5 min for 2.5 h. Focus was maintained by the Perfect Focus System (Nikon).

Nuclear actin-chromobody TagGFP imaging and analysis. For live cell imaging experiments, U2OS cells were cultured on 35-mm glass bottom microwell dishes at 90% confluency (MatTek, P35GC-1.5-10-C) and transfected with 0.3 μg of vector expressing nuclear actin-cb (Chromotek, acg-n; EMD Millipore, FuGENE 6 transfection reagent; Promega) After 6 h, cells were treated with 1 μg/ml NCS for 60 min at 37 °C. Cells were washed twice with PBS before the addition of medium containing DMSO. After 3 h, images were acquired on an A1RMP confocal microscope using the equipment and settings described above. Cells with actin-cb foci were classified using the following criteria: actin-cb foci exhibited movement and frequent turnover, were round in morphology and often appeared to have fuzzy 'tails'. Cells with actin-cb rods were classified using the following criteria: actin-cb rods exhibited limited movement, no turnover, were very bright, and were filamentous in morphology. Foci movements were examined by collecting *z* series at 0.4-μm intervals throughout the entire nucleus every 30 s for 10 min. Following initial image acquisition, cells were subsequently treated with 100 μM CK-666. Images of actin-cb foci were subsequently acquired 1 h post CK-666 treatment. Actin-cb foci were scored as the number of discrete punctate structures appearing per nucleus, as detected by the Icy Spot Detector plug-in. For experiments assessing localization of actin-cb foci with RPA, U2OS cells were co-transfected with actin-cb and RPA32–NLS–mCherry (a gift from J. Lukas) for 16 h before NCS treatment. After 12 h, foci movements were examined by collecting *z* series at 0.4-μm intervals throughout the entire nucleus every 5 min for 2.5 h.

For experiments in fixed cells, U2OS cells were cultured on 8-well chamber slides (Thermo Fisher Scientific), subjected to transfection and drug treatments as described above, and processed for IHC as described above. Cells were stained with primary antibodies against RAD51 (Santa Cruz: sc-8349, 1/50) and GFP (Abcam: ab13970, 1/100) overnight. Secondary antibodies were Alexa 488 conjugated goat anti-chicken Ig (Invitrogen: A11039, 1/1000) and Alexa 594 conjugated goat anti-mouse Ig (Thermo Fisher Scientific, A-11005, 1/10000). Actin-cb-expressing cells were selected for analysis by screening for actin-cb foci. Images were acquired on an A1RMP confocal microscope equipped with a 60×/1.49 Apo-TIRF oil-immersion objective lens. *Z* series at 0.4-μm intervals throughout the entire nucleus were acquired. For co-localization analysis, actin-cb and Rad51 foci were detected in *z* using the spot detector plug-in available on Icy Bioimaging Software platform. The co-localization of RAD51 and actin-cb spots was assessed using the Co-localization Studio plug-in, object-based methodology. Here, co-localization was defined as the number of RAD51 spots that directly overlapped with or touched actin-cb foci (within a 0.7-μm radius).

Analysis of movement of DSB foci. DSB movement analysis followed the approach recently described[9]. Nikon NIS Elements data files were processed in ImageJ[39]. Green and red channels were split and a maximum-intensity projection of each *z* stack was generated. T stacks were subsequently aligned using the StackReg plugin[40] to correct for cell movements over the duration of the experiment. Cells that underwent large-scale nuclear deformations or expansions were discarded. The TrackMate plug in for ImageJ[41] was used to perform single-particle tracking of DSB foci over a 100-min interval and monitor foci intensity. Focus trajectories were subsequently exported to MATLAB and analysed using the class @msdanalyzer[42]. The MSD of DNA damage foci plots the average squared distance

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

travelled by foci at increasing time intervals, whereas the diffusion coefficient $D(t)$ is approximated through the linear-weighted fit of the initial mean MSD curve[42]. MSD curves for DSB foci in a given cell were computed using the formula $\mathrm{MSD} = (x(t+\Delta t) - x(t))^2$, where x reflects focus position and t is the time in minutes. For RAD52 foci, the weighted mean of all the MSD curves acquired from >2,000 tracks in three independent experiments is shown. The error bars for each point on the weighted mean of the MSD curve represent the weighted s.e.m. over all MSD curves. The diffusion coefficient $D(t)$ was estimated from the linear fit of the first 20% of each MSD curve. Calculation of the time-dependence coefficient α was performed using MATLAB via log–log fitting of the power law $\mathrm{MSD}(t) = \Gamma \times t^{\alpha}$. Cumulative distance (CD) was calculated by summing the distances foci travelled over 100 min relative to their starting position using the formula: $D(i) = \sqrt{(X(i) - X(i-1))^2 + (Y(i) - Y(i-1))^2}$ where X and Y refer to the x and y coordinates of the focus at time i.

Analysis of clustering of DSB foci. For analysis of clustering of DSB foci during live cell imaging, a clustering event was defined as complete co-localization of ≥2 foci over ≥3 consecutive frames (15 min) over a 100-min interval. For actin-cb clustering during live cell imaging, a clustering event was defined as complete co-localization of ≥2 foci over ≥3 consecutive frames (3 min) over a 20-min interval.

Quantification of RPA association to chromatin by flow cytometry. Detection of chromatin-bound RPA in S-phase cells followed recently published protocols with the following modifications: 2×10^6 cells were treated with DMSO or 1 μg/ml CPT for 24 h. Cells were subsequently washed once with cold PBS, then incubated in cold CSK buffer + 1% Triton-X for 5 min. Cells were then washed with PBS and fixed in 0.5% PFA for 15 min. Cells were subsequently washed twice in BD Perm/Wash buffer (Becton Dickinson: BD554723) and stained with RPA2/RPA32 antibody (Abcam: ab2175, 1/50) overnight. Cells were then washed with PBS and incubated in secondary antibody (Abcam: ab150113, 1/50 or Alexa 647 conjugated donkey anti-mouse Ig, Invitrogen: A31571, 1/50) for 1 h. Cells were then washed and stained with propidium iodide (PI) in the presence of RNase A overnight. The percentage of S-phase cells as identified by PI staining was quantified by flow cytometry. The gate for RPA-positive cells was established using a negative control sample that was stained with mouse IgG following extraction. RPA-positive cells following DMSO and CPT treatment were subsequently quantified by flow cytometry. A minimum of 10,000 cells was counted per biological replicate.

Clonogenic and cell survival assays. U2OS cells were seeded onto 10-cm dishes (1,000 cells per dish) overnight. Cells were subsequently treated with 0, 100, 200, or 300 μM CK-666 in the presence of DMSO, CPT (10 nM), or aphidocolin (APH, 400 nM) for 12 h. Cells were then washed with PBS and incubated in fresh medium for 10–14 days. Cells were subsequently washed with PBS, fixed for 10 min with 100% methanol, and stained with 0.1% crystal violet for 10 min. For olaparib studies, cells were treated with 0, 10, or 20 μM CK-666 in the presence of DMSO or 50 nM olaparib for 14 days, and fixed and stained as above. Each experiment was performed in triplicate.

For colony quantification, images of dishes were converted to greyscale and processed with Icy Software (Version 1.8.6.0; Institut Pasteur, Quantitative Image Analysis Unit; http://www.icy.bioimageanalysis.org). Regions of interest were drawn around individual dishes and colonies were counted using the Spot Detector plug-in (Channel 0). Counts from two biological replicates were normalized for plating efficiency.

For the annexin V/PI survival assay, $1–5 \times 10^5$ cells were incubated in DMSO or 1 μg/ml CPT for 0, 12, or 24 h. Cells were then washed in 1× binding buffer (Thermo Fisher Scientific: 00-0055) and stained with APC-conjugated annexin V (Thermo Fisher Scientific: 17-8007) for 10 min. Cells were then washed in 1× binding buffer and stained in PI (Thermo Fisher Scientific: 00-6990). Per cent viability following CPT treatment was assessed by measuring the percentage of APC-Annexin V-negative and PI-negative cells by flow cytometry. A minimum of 10,000 cells counted per condition.

Quantification and statistical analysis. Statistical parameters are reported in the figures and legends. Two-tailed t-tests and ordinary one-way ANOVAs were used for comparisons of means of normally distributed data and two-tailed Mann–Whitney U tests were used for comparison of non-normally distributed data. Statistical analyses were performed using Prism 7.0. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Research animals. The use of *Xenopus laevis* (African clawed frog) in this research proposal was approved (protocol AAAK0551) by the Institutional Animal Care and Use Committee (IACUC) of Columbia University Medical School. All procedures comply with Public Health Service's policies for the humane care and use of laboratory animals.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

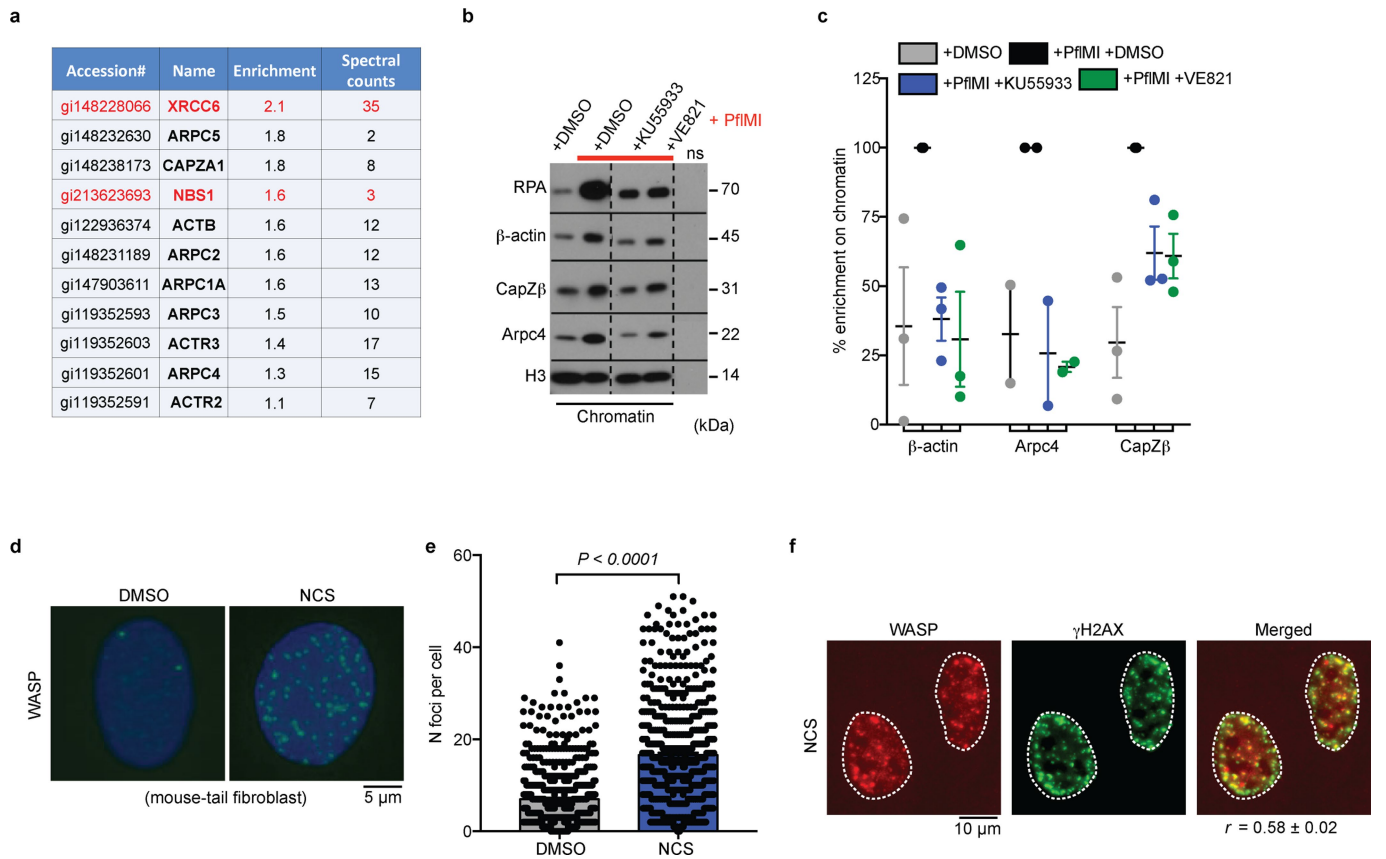
Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request. The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information files.

36. Peterson, S. E. et al. Cdk1 uncouples CtIP-dependent resection and Rad51 filament formation during M-phase double-strand break repair. *J. Cell Biol.* **194**, 705–720 (2011).
37. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
38. Lagache, T., Lang, G., Sauvonnet, N. & Olivo-Marin, J. C. Analysis of the spatial organization of molecules with robust statistics. *PLoS ONE* **8**, e80914 (2013).
39. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
40. Thevenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans Image Process* **7**, 27–41 (1998).
41. Tinevez, J. Y. et al. TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 (2017).
42. Tarantino, N. et al. TNF and IL-1 exhibit distinct ubiquitin requirements for inducing NEMO-IKK supramolecular structures. *J. Cell Biol.* **204**, 231–245 (2014).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

a

Accession#	Name	Enrichment	Spectral counts
gi148228066	XRCC6	2.1	35
gi148232630	ARPC5	1.8	2
gi148238173	CAPZA1	1.8	8
gi213623693	NBS1	1.6	3
gi122936374	ACTB	1.6	12
gi148231189	ARPC2	1.6	12
gi147903611	ARPC1A	1.6	13
gi119352593	ARPC3	1.5	10
gi119352603	ACTR3	1.4	17
gi119352601	ARPC4	1.3	15
gi119352591	ACTR2	1.1	7



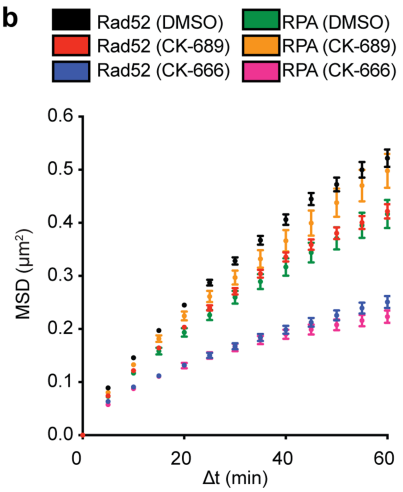
Extended Data Fig. 1 | Actin filament nucleators localize to chromosomal DSBs in *Xenopus* extracts and mammalian cells. **a**, Enrichment ratios and spectral counts of actin complexes and repair proteins in DSB-containing chromatin are shown by liquid chromatography–mass spectrometry. **b**, DNA damage-dependent enrichment of actin complexes (+PflMI) following PIKK inhibition (KU55933: ATMi; VE821: ATRi). RPA shown as a marker of DNA damage. **c**, Quantification of actin complexes in chromatin relative to +PflMI samples. β-actin $n = 3$ independent experiments, ARPC4 $n = 2$ independent experiments, CAPZβ $n = 3$ independent experiments. Mean and s.e.m shown. **d**, Representative images of WASP foci after NCS treatment in MTFs. **e**, Quantification of WASP foci (P calculated by two-sided Mann–Whitney test; data shown as mean; DMSO $n = 578$ nuclei, NCS $n = 556$ nuclei). **f**, Co-localization of γH2AX with WASP foci in MTFs. $n = 22$ nuclei, $r = 0.58 \pm 0.021$, Pearson. Scale bar, 5 μm or as indicated.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

a

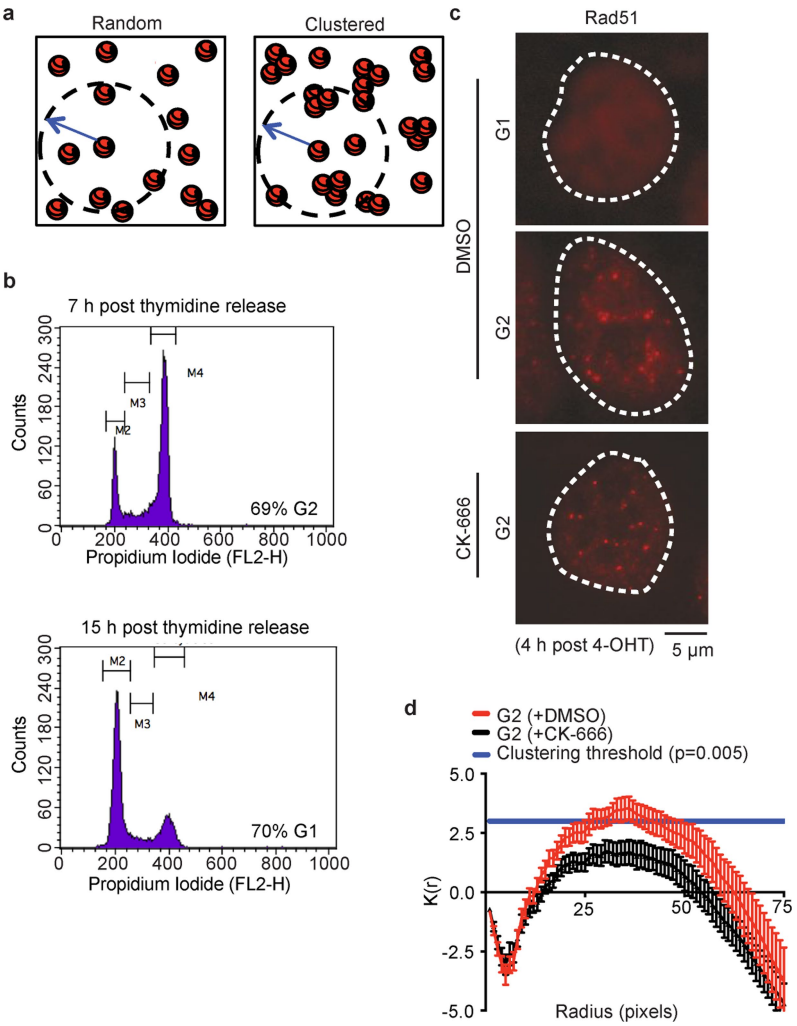
	Treatment	No cells analyzed (No foci)	Cumulative Distance ± SD (μm)	Diffusion coefficient (10^{-5} μm^2/s)	Anomalous Diffusion Coefficient ± SE	No of experiments
Rad52-mCherry, U2OS (100 min)	DMSO	12 (3292)	4.2 ± 1.29	4.38	0.719 ± 0.006	3
	CK-689 (100 μM)	13 (3262)	3.8 ± 1.08	3.65	0.726 ± 0.006	3
	CK-666 (100 μM)	12 (2143)	3.2 ± 1.05	1.96	0.589 ± 0.009	3
	Mirin (50 μM)	11 (2677)	3.3 ± 0.92	2.36	0.643 ± 0.007	2
53BP1-YFP, U2OS (G1) (100 min)	DMSO	12 (893)	3.6 ± 1.03	2.93	0.715 ± 0.011	2
	CK-689 (100 μM)	14 (926)	3.3 ± 0.89	2.41	0.709 ± 0.012	2
	CK-666 (100 μM)	14 (1234)	3.2 ± 1.03	2.32	0.669 ± 0.009	2
	Mirin (50 μM)	14 (774)	3.3 ± 1.05	2.35	0.685 ± 0.013	2
RPA32-pEGFP-NLS, MTF (100 min)	DMSO	10 (1031)	3.7 ± 1.03	3.27	0.731 ± 0.013	2
	CK-689 (100 μM)	13 (790)	4.0 ± 1.25	4.08	0.743 ± 0.012	2
	CK-666 (100 μM)	12 (823)	3.4 ± 1.01	2.07	0.66 ± 0.013	2
	Arpc2 -/- (4μM 4-OHT)	13 (1135)	3.2 ± 1.09	1.76	0.60 ± 0.011	2

b



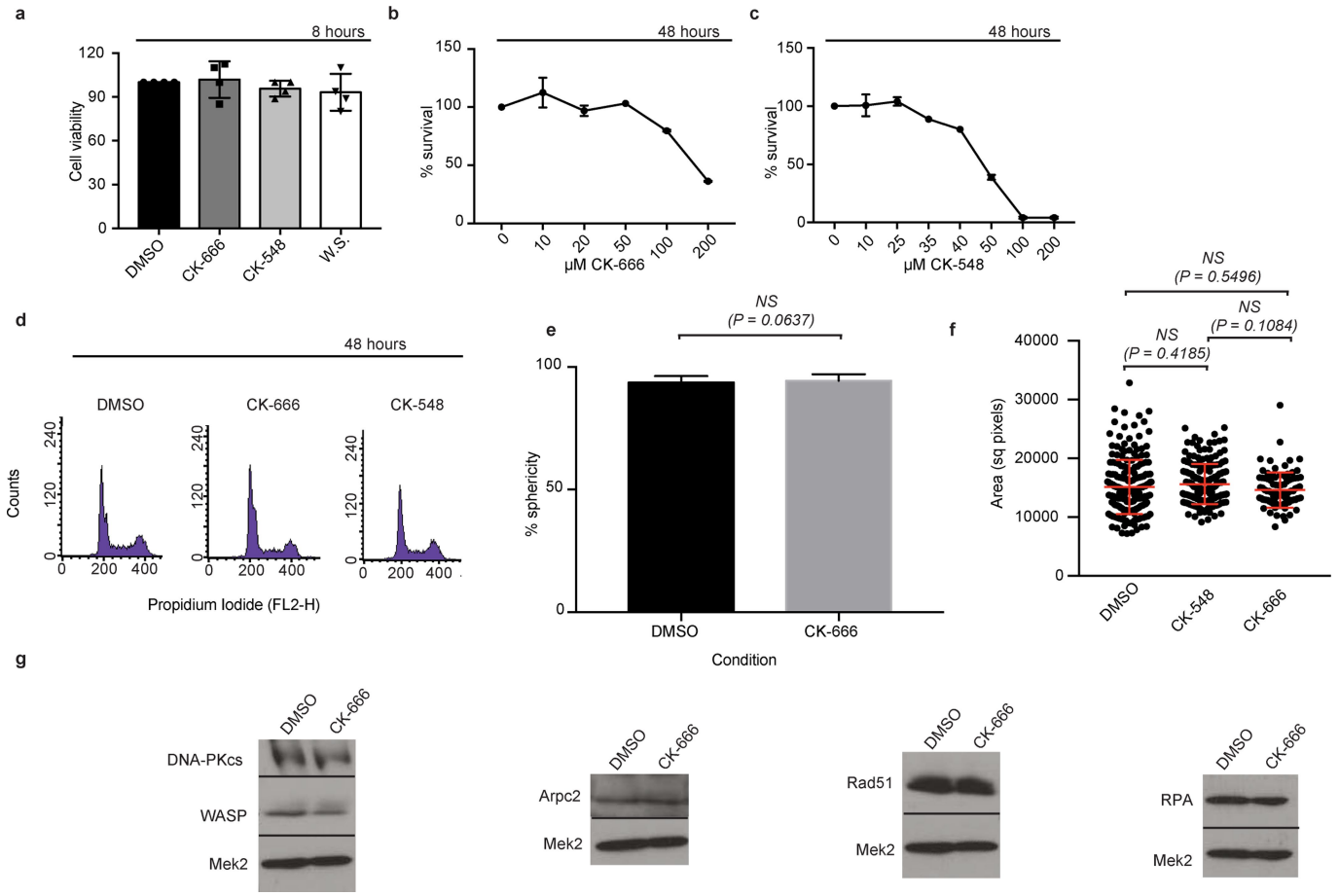
Extended Data Fig. 2 | CK-689 does not significantly alter movement of foci. a, Summary of DSB movement analyses as related to Figs. 3, 6, and Extended Data Fig. 5. **b**, MSD of RAD52–mCherry foci and RPA–pEGFP–NLS foci treated with DMSO, CK-689 and CK-666. Data shown as mean and weighted s.e.m.; RAD52: CK-689 $n = 3,262$ foci from 13 nuclei, CK-666 $n = 2,143$ foci from 12 nuclei, DMSO $n = 3,292$ foci from 12 nuclei; RPA32: CK-689 $n = 790$ foci from 13 nuclei, CK-666 $n = 823$ foci from 12 nuclei, DMSO $n = 1,031$ foci from 10 nuclei. MSD of RAD52 DMSO also shown in Fig. 6i. MSD of RAD52 CK-689 and CK-666 also shown in Fig. 3g. MSD of RPA CK-689 and CK-666 also shown in Fig. 3k.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



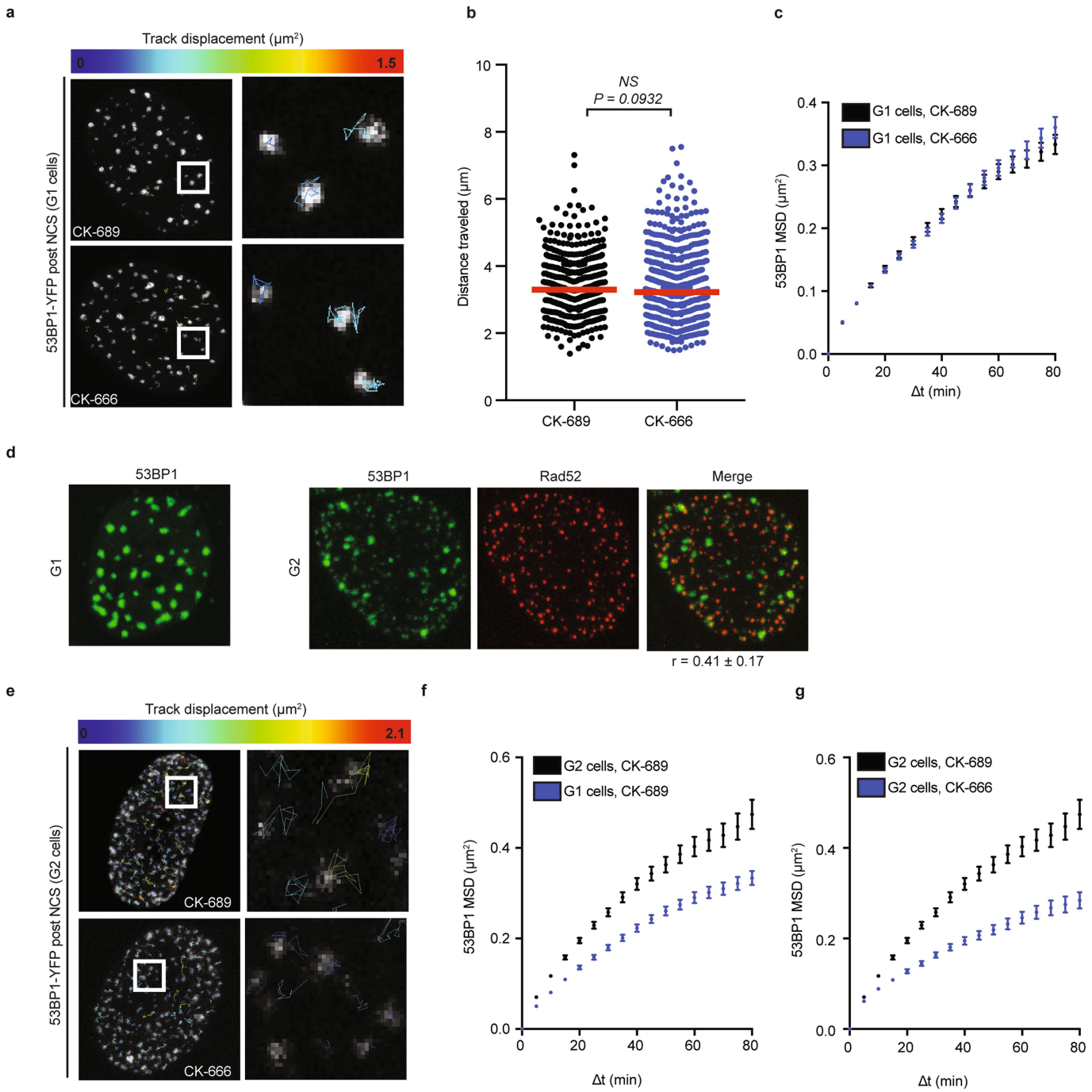
Extended Data Fig. 3 | ARP2/3 clusters RAD51 foci. a, Schematic of clustering events, as defined by the number of paired points that fall within a given radius (arrow, dashed circle). **b**, Representative cell cycle distribution of U2OS cells following double-thymidine block (out of five independent experiments). **c**, Representative images of U2OS-AsiSI nuclei showing RAD51 foci. **d**, RAD51 foci clustering in G2 cells (*P* calculated by Spatial analysis plug-in; data shown as mean and s.e.m. of Ripley function; DMSO $n = 95$ nuclei, CK-666 $n = 80$ nuclei).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



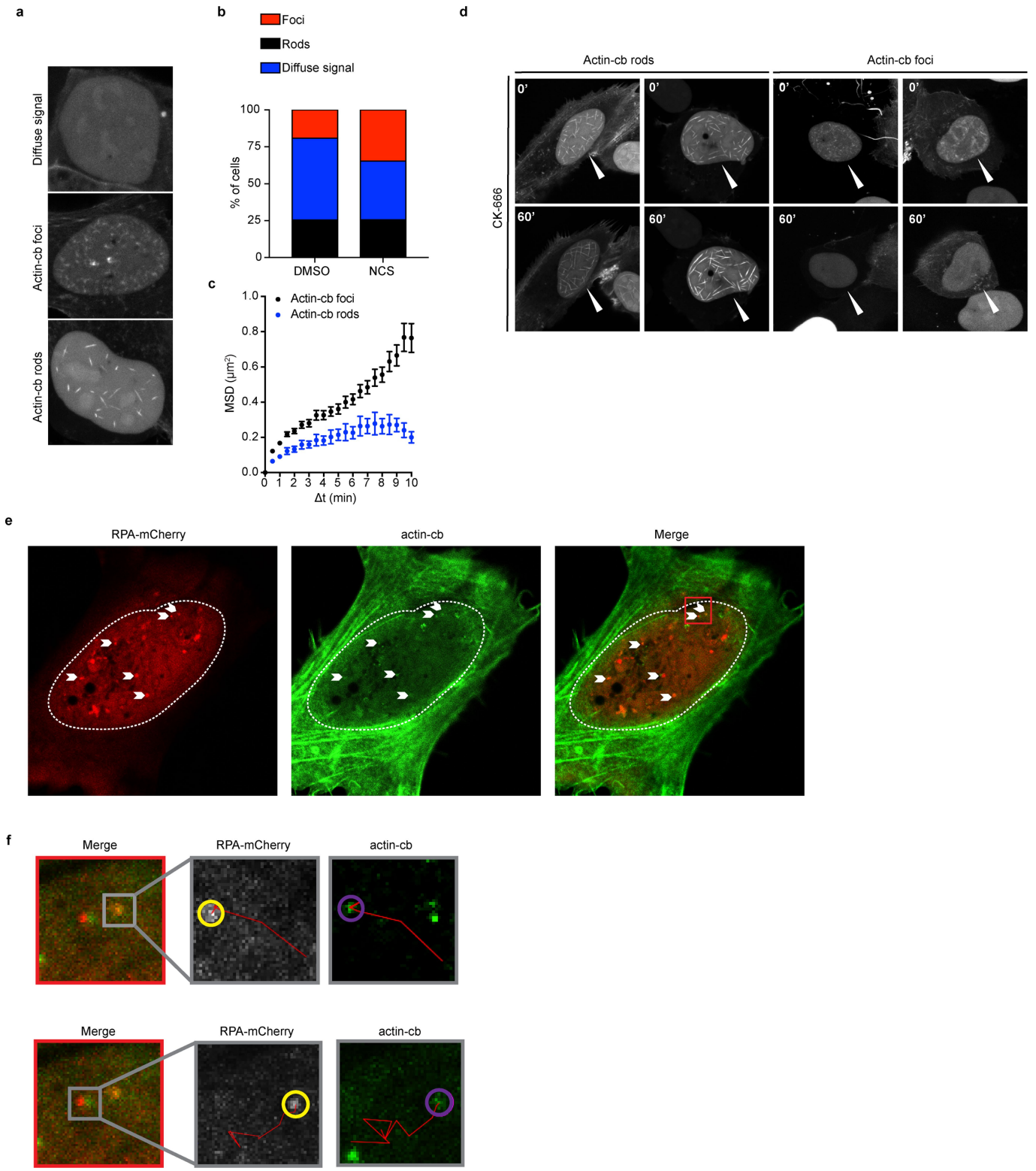
Extended Data Fig. 4 | WASP and ARP2/3 inactivation do not impair U2OS cell viability, nuclear area, nuclear sphericity or protein expression. **a**, Cell survival after 8 h of treatment with 100 μM CK-666, 50 μM CK-548, or 3 μM wiskostatin. Data shown as mean and s.e.m. $n = 4$ independent experiments. **b**, Cell survival after 48 h exposure to increasing concentrations of CK-666. $n = 3$ independent experiments. Data shown as mean and s.e.m. **c**, Cell survival after 48 h exposure to increasing concentrations of CK-548. $n = 3$ independent experiments. Data shown as mean and s.e.m. **d**, Representative cell-cycle distribution of cells treated with DMSO, 50 μM CK-666, or 25 μM CK-548 for 48 h (out of three independent experiments). **e**, Analysis of nuclear sphericity. *P* calculated by Student's two-tailed *t*-test; data shown as mean and s.d.; DMSO $n = 117$ nuclei, CK-666 $n = 117$ nuclei. NS, not significant. **f**, Analysis of nuclear area. *P* calculated by one-way ANOVA with multiple comparisons; data shown as mean and s.d.; DMSO $n = 210$ nuclei, CK-548 $n = 189$ nuclei, CK-666 $n = 92$ nuclei. **g**, Expression of RAD51, DNA-PKcs, WASP, and ARPC2 validating antibodies used in Fig. 2e–i (single experiment). MEK2 is a loading control. RPA expression levels also shown.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



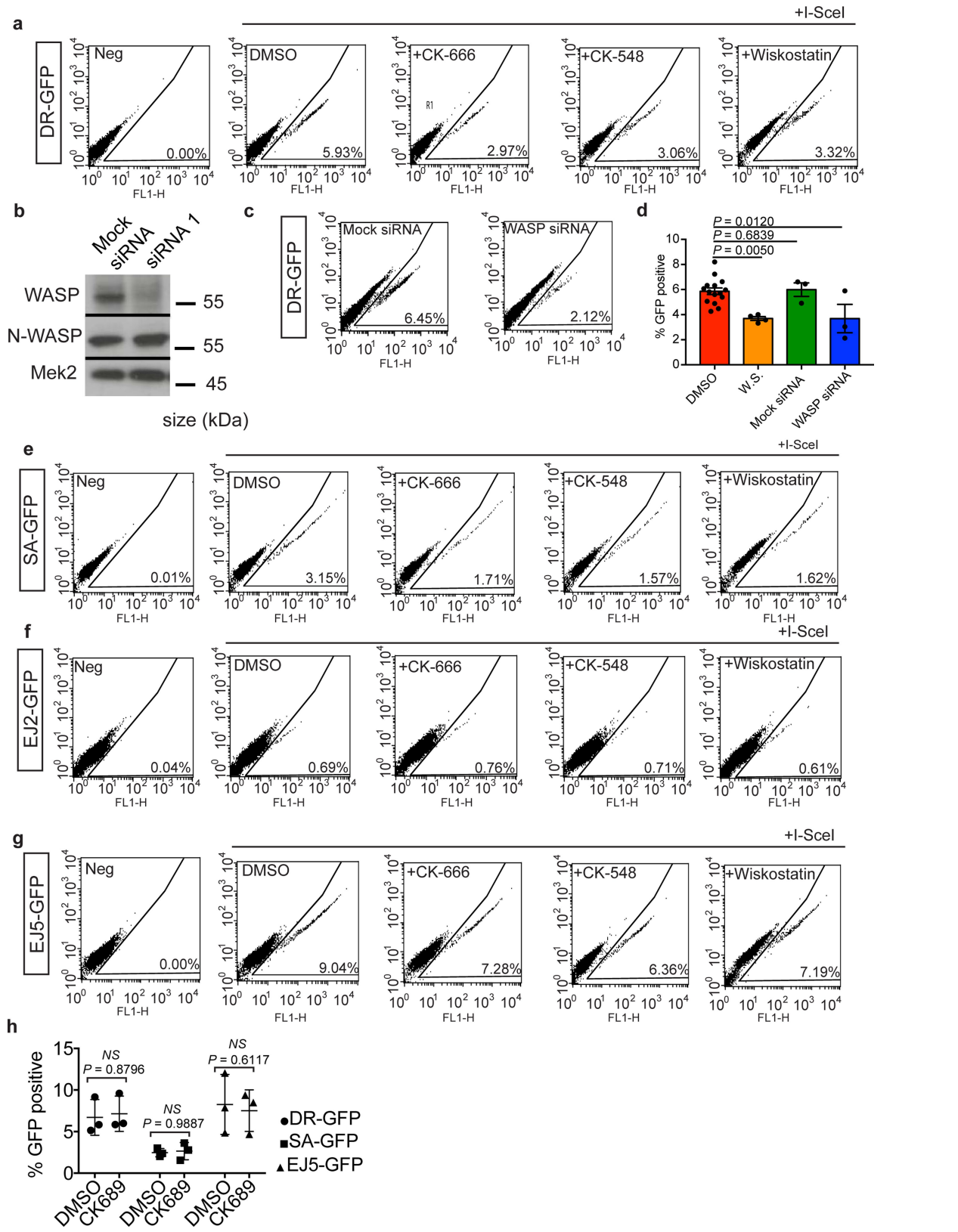
Extended Data Fig. 5 | ARP2/3 enhances movement of 53BP1 foci in G2 cells. a, Representative U2OS nuclei showing 53BP1–YFP focus traces over 100 min in G1 cells. **b**, Median cumulative distance travelled by 53BP1–YFP foci in G1 cells. *P* calculated by two-tailed Mann–Whitney test; CK-689 $n = 462$ foci from 14 nuclei, CK-666 $n = 647$ foci from 14 nuclei. NS, not significant. **c**, MSD of 53BP1–YFP foci in G1 cells. Data shown as mean and weighted s.e.m.; CK-689 $n = 926$ foci from 14 nuclei, CK-666 $n = 1{,}234$ foci from 14 nuclei. Δt, time interval. **d**, Left, a representative G1 cell with 53BP1–YFP foci. Right, a representative G2 cell with 53BP1–YFP and Rad52–mCherry foci. 53BP1 foci colocalize with RAD52 foci. $r = 0.41 \pm 0.17$, Pearson, $n = 5$ independent experiments. **e**, Representative U2OS nuclei showing traces of 53BP1–YFP foci over 100 min in G2 cells. **f**, MSD of 53BP1–YFP foci. Data shown as mean and weighted s.e.m.; G1 CK-689 $n = 926$ foci from 14 nuclei, G2 CK-689 $n = 1{,}403$ foci from 12 nuclei. G1 CK-689 curve also shown in **c**. **g**, MSD of 53BP1–YFP foci in G2 cells. Data shown as mean and weighted s.e.m.; CK-689 $n = 1{,}403$ foci from 12 nuclei, CK-666 $n = 1{,}038$ foci from 10 nuclei. G2 CK-689 curve also shown in **f**.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



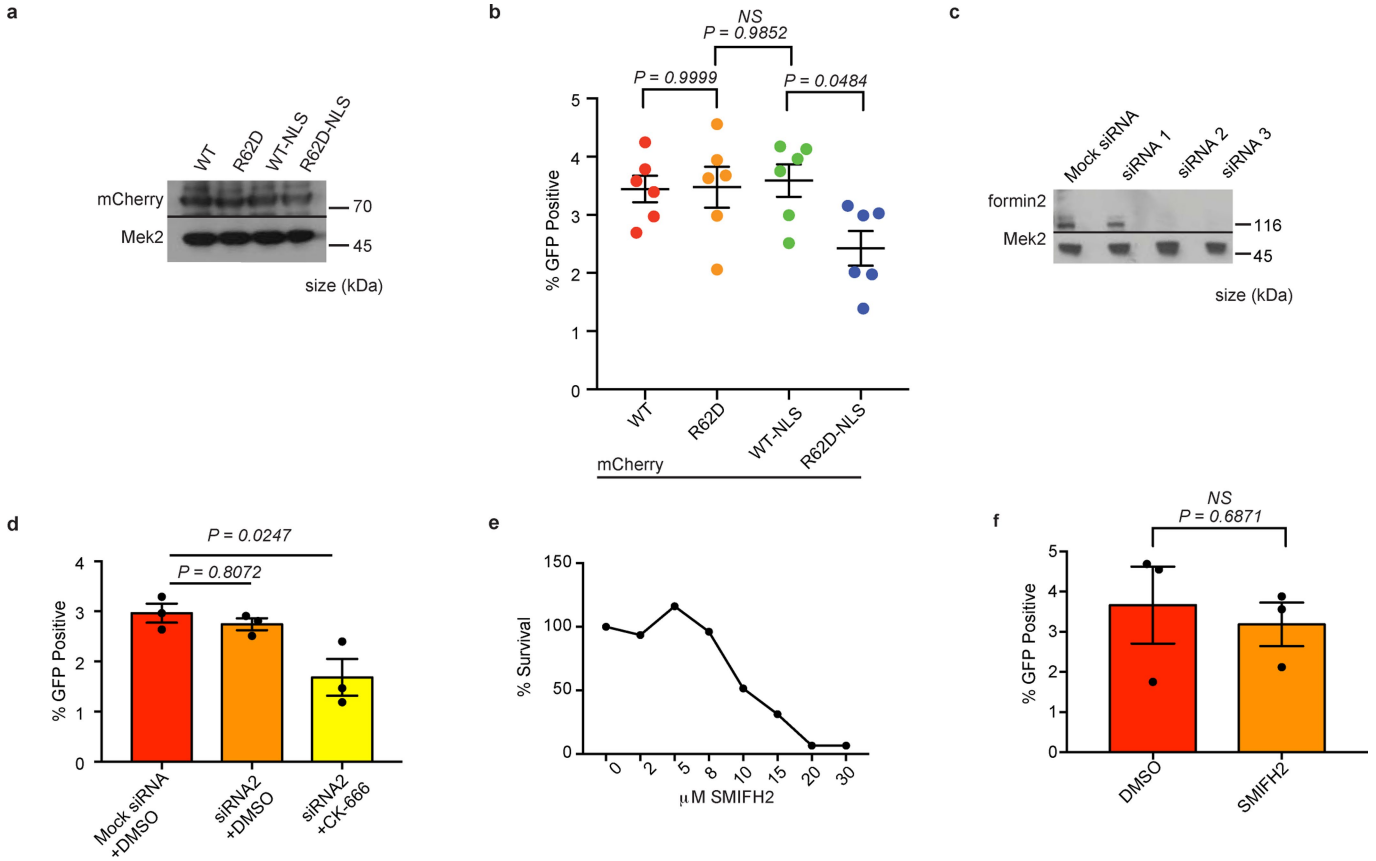
Extended Data Fig. 6 | ARP2/3 promotes assembly of actin foci following DNA damage. a, Representative U2OS nuclei showing classes of nuclear actin structures following transient transfection with nuclear actin-chromobody TagGFP. **b**, Percentage of cells with diffuse signal, nuclear actin-cb foci, or rods with or without NCS (DMSO $n = 473$ cells, NCS $n = 473$ cells). **c**, MSD of actin-cb foci and actin-cb rods. Data shown as mean and weighted s.e.m.; actin-cb foci $n = 662$ foci from 11 nuclei, actin-cb rods $n = 161$ rods from 5 nuclei. **d**, Representative images of U2OS nuclei with actin rods or actin-cb foci following CK-666 treatment ($n = 5$ independent experiments). **e**, Representative image of a U2OS nucleus with RPA–mCherry and actin-cb foci of $n = 3$ independent experiments. Arrowheads indicate sites of RPA–mCherry and actin-cb co-localization. **f**, Expanded image of RPA–mCherry and actin-cb co-localization from **e** (red box). Traces of RPA and actin-cb foci are shown in red. Yellow and purple circles encompass RPA and actin-cb foci, respectively. $n = 3$ independent experiments.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



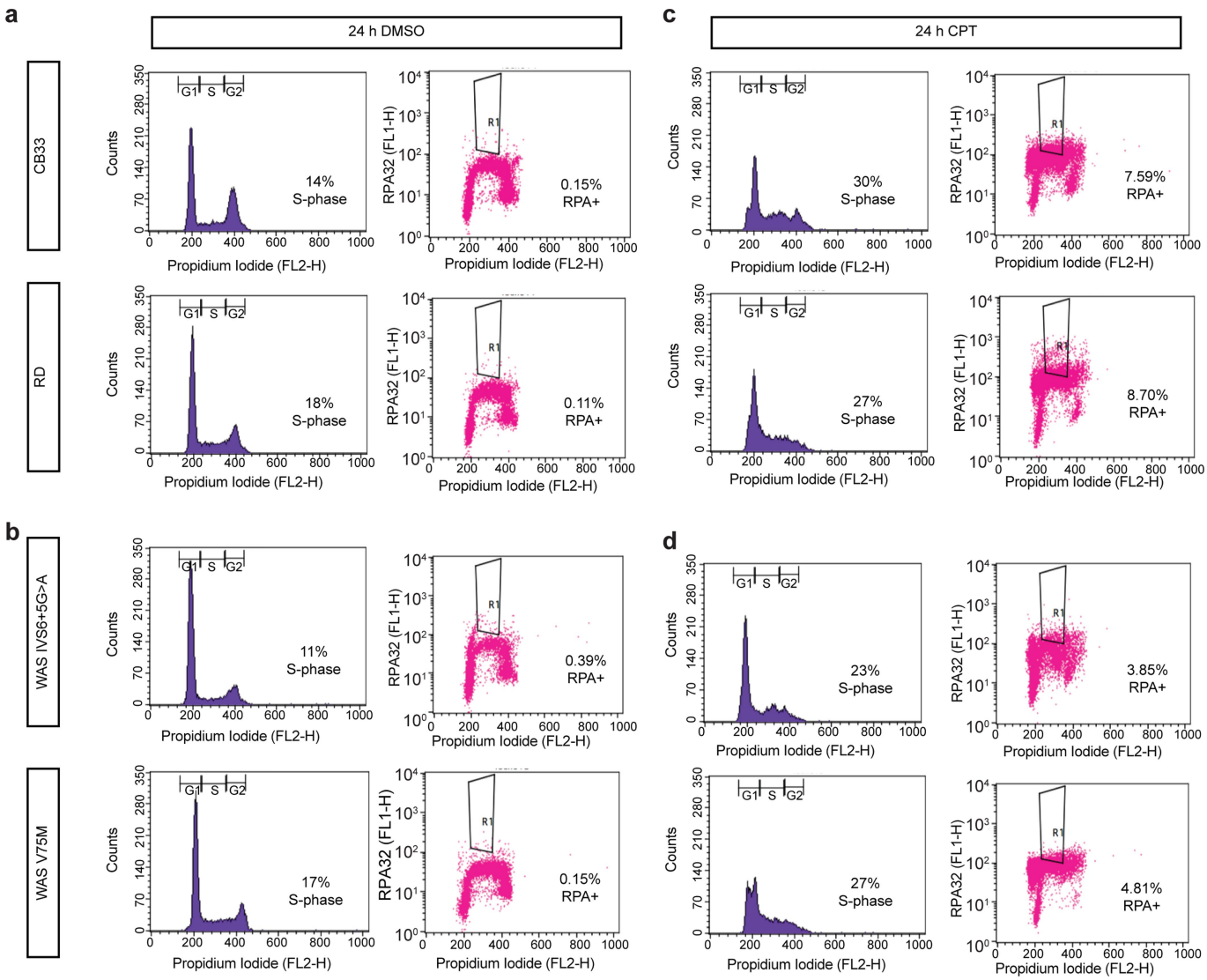
Extended Data Fig. 7 | WASP and ARP2/3 mediate DSB repair by homology-directed mechanisms a, Representative FACS plots of GFP$^+$ cells in the HDR (DR-GFP) assay. **b**, Western blot shows expression of WASP and N-WASP following WASP siRNA knockdown in whole-cell lysates (single experiment). WASP siRNA 1: 5′-GAGUGGCUGAGUUACUUGC-3′. **c**, Representative FACS plots of GFP$^+$ cells in the HDR (DR-GFP) assay in WASP and mock-depleted cells. **d**, Summary of DR-GFP assay with WASP depletion. *P* calculated by one-way ANOVA with multiple comparisons; data shown as mean and s.e.m; $n = 3$ independent experiments. HDR efficiency in the presence of DMSO or wiskostatin (Fig. 5a) shown for comparison. **e**, Representative FACS plots of GFP$^+$ cells in the SSA (SA-GFP) assay. **f**, Representative FACS plots of GFP$^+$ cells in the MMEJ (EJ2-GFP) assay. **g**, Representative FACS plots of GFP$^+$ cells in the NHEJ (EJ5-GFP) assay. **h**, Summary of DR-GFP, SA-GFP, and EJ5-GFP assays with CK-689. *P* calculated by two-way ANOVA with multiple comparisons; data shown as mean and s.d.; $n = 3$ independent experiments. NS, not significant.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



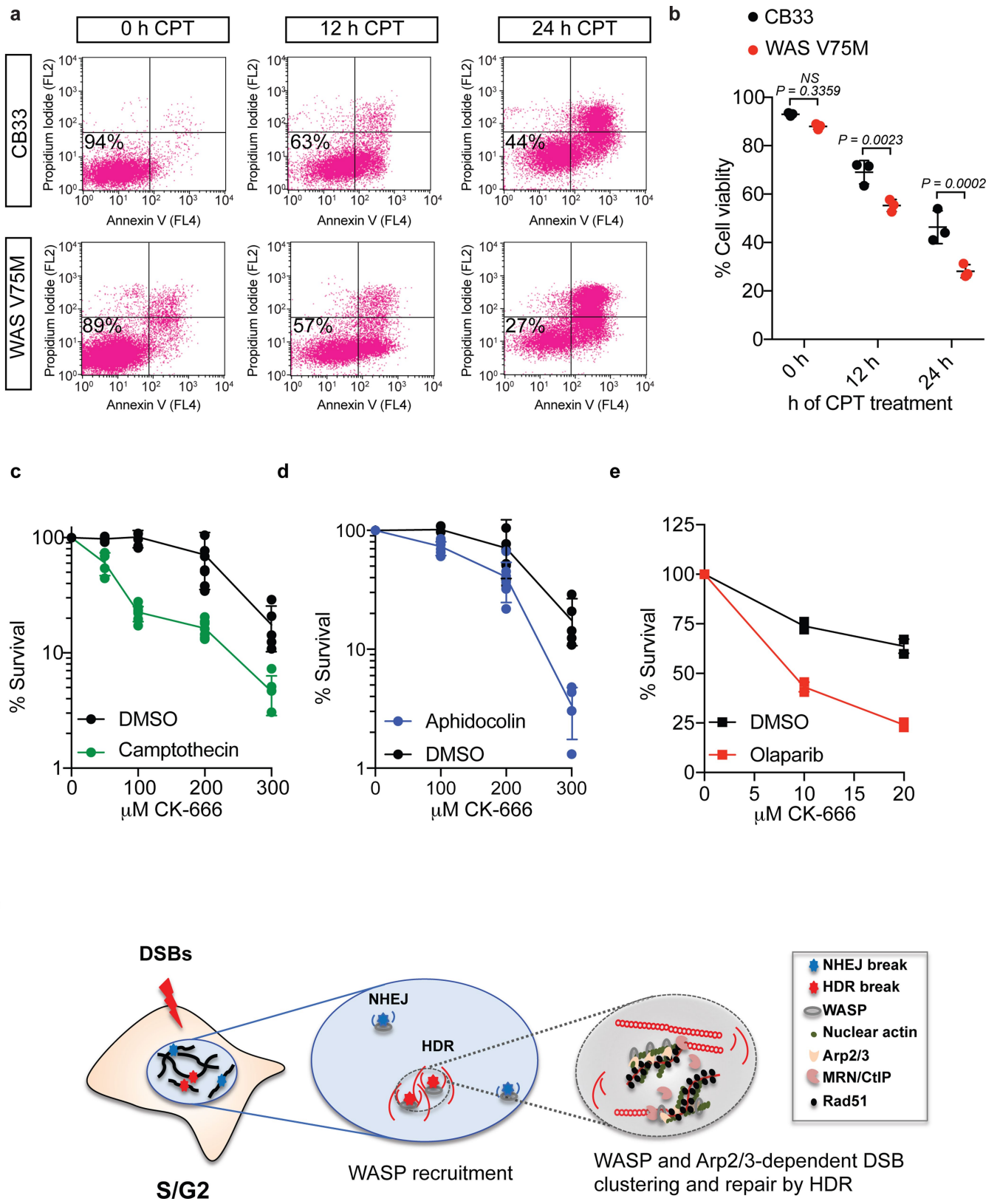
Extended Data Fig. 8 | Actin nucleation regulates HDR in the nucleus and does not require formin-2 activity. a, Western blot shows expression of mCherry-tagged actin constructs in U2OS whole-cell lysates (single experiment). **b**, Summary of DR-GFP (HDR) assay. *P* calculated by one-way ANOVA with multiple comparisons; data shown as mean and s.e.m; $n = 6$ independent experiments. **c**, Western blot shows knockdown of formin-2 following 48 h transfection of mock or sequence-specific siRNA oligos (single experiment). Formin-2 siRNA #2: 5′-CGUGUAAUCAGAAUGCCCA-3′. **d**, Summary of DR-GFP assay. *P* calculated by one-way ANOVA with multiple comparisons; data shown as mean and s.e.m; $n = 3$ independent experiments. **e**, Mean cell survival after 48 h exposure to increasing concentrations of the formin inhibitor SMIFH2 ($n = 3$ independent experiments). **f**, Summary of DR-GFP assay. *P* calculated by Student's two-tailed *t*-test; data shown as mean and s.e.m; $n = 3$ independent experiments.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 9 | B lymphocytes derived from patients with WAS exhibit reduced DSB end-resection. a, Representative cell cycle distribution of CB33 and RD lymphocytes derived from healthy individuals following DMSO treatment (left). The percentage of RPA-positive S-phase cells following DMSO treatment was measured by flow cytometry (right). **b**, Representative cell cycle distribution of lymphocytes bearing the V75M mutation in the *WAS* gene or a G>A transition at position 5 in intron 6 of the *WAS* gene (IVS6+5G>A) following DMSO treatment (left). The percentage of RPA-positive S-phase cells following DMSO treatment was measured by flow cytometry (right). **c**, Representative cell cycle distribution of CB33 and RD lymphocytes following CPT treatment (left). The percentage of RPA-positive S-phase cells following CPT treatment was measured by flow cytometry (right). **d**, The cell cycle distribution of *WAS* V75M or *WAS* IVS6+5G>A lymphocytes following CPT treatment (left). The percentage of RPA-positive S-phase cells following CPT treatment was measured by flow cytometry (right). For all panels, $n = 4$ independent experiments.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 10 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 10 | ARP2/3 inactivation confers sensitivity to DSBs induced in S-phase as well as replication-stress-inducing agents. **a**, Control CB33 lymphocytes or lymphocytes bearing a V75M mutation in the *WAS* gene were treated with CPT for 0, 12, or 24 h. Per cent viability following CPT treatment was assessed by measuring the fraction of annexin V- and PI-negative cells by flow cytometry. **b**, Summary of CB33 or *WAS* V75M lymphocyte survival following CPT treatment. *P* calculated by two-way ANOVA with multiple comparisons; data shown as mean and s.d.; $n = 3$ independent experiments. **c**, Clonogenic U2OS cell survival after 12 h of CPT treatment in the presence of DMSO or increasing concentrations of CK-666. Triplicate experiments; data shown as mean and s.d.; $n = 2$ independent experiments. **d**, Clonogenic U2OS cell survival after 12 h of aphidicolin treatment in the presence of DMSO or increasing concentrations of CK-666. Triplicate experiments; data shown as mean and s.d.; $n = 2$ independent experiments. **e**, Clonogenic U2OS cell survival after olaparib treatment in the presence of DMSO or increasing concentrations of CK-666 for 14 days. Triplicate experiments; data shown as mean and s.e.m.; $n = 2$ independent experiments. **f**, DNA damage induces DSBs, which are repaired preferentially by NHEJ in mammalian cells (blue). In S/G2, DSBs may be repaired by HDR (red). All DSBs recruit WASP, but ARP2/3-dependent actin polymerization occurs only at HDR breaks, which become more mobile. Actin polymerization in the vicinity of DSBs generates forces that result in DSB clustering, optimal DNA end resection, formation of RAD51 foci, and HDR.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ARTICLE

https://doi.org/10.1038/s41586-018-0255-3

Structure of a human synaptic $GABA_A$ receptor

Shaotong Zhu[1], Colleen M. Noviello[1], Jinfeng Teng[1], Richard M. Walsh Jr[1], Jeong Joo Kim[1] & Ryan E. Hibbs[1]*

Fast inhibitory neurotransmission in the brain is principally mediated by the neurotransmitter GABA (γ-aminobutyric acid) and its synaptic target, the type A GABA receptor ($GABA_A$ receptor). Dysfunction of this receptor results in neurological disorders and mental illnesses including epilepsy, anxiety and insomnia. The $GABA_A$ receptor is also a prolific target for therapeutic, illicit and recreational drugs, including benzodiazepines, barbiturates, anaesthetics and ethanol. Here we present high-resolution cryo-electron microscopy structures of the human α1β2γ2 $GABA_A$ receptor, the predominant isoform in the adult brain, in complex with GABA and the benzodiazepine site antagonist flumazenil, the first-line clinical treatment for benzodiazepine overdose. The receptor architecture reveals unique heteromeric interactions for this important class of inhibitory neurotransmitter receptor. This work provides a template for understanding receptor modulation by GABA and benzodiazepines, and will assist rational approaches to therapeutic targeting of this receptor for neurological disorders and mental illness.

The function of the nervous system is governed by a balance of excitatory and inhibitory signalling. GABA is the major inhibitory neurotransmitter in the central nervous system (CNS) and acts through the $GABA_A$ and $GABA_B$ receptors. $GABA_A$ receptors, found at 20–50% of synapses in the brain[1], react on a millisecond timescale to binding of GABA by opening a transmembrane channel that is permeable to chloride, which suppresses neuronal activity in the adult brain[2]. Dysfunction of these channels results in anxiety disorders, epilepsy, and neurodevelopmental disorders, including autism[3–5].

$GABA_A$ receptors are the targets of a remarkably diverse array of drugs that act through distinct binding sites. GABA was discovered in 1950[6,7], and was soon followed by the discovery of benzodiazepines[8], allosteric modulators of $GABA_A$ receptors that are widely used in the treatment of epilepsy, insomnia, anxiety and panic disorder[9,10]. Flumazenil is a competitive antagonist of the benzodiazepine-binding site; it is used clinically to reverse benzodiazepine-induced anaesthesia and is the principal antidote for benzodiazepine overdose[11]. Allosteric potentiation of the $GABA_A$ receptor for therapeutic (or recreational) purposes extends far beyond benzodiazepines: barbiturates, volatile and intravenous anesthetics, neurosteroids and ethanol are all allosteric modulators that act on $GABA_A$ receptors[12,13].

The rich pharmacology of the $GABA_A$ receptor derives in part from its complex subunit assembly. A total of 19 subunits assemble in limited combinations to make functional receptors[14]. The predominant synaptic isoform comprises two α1-subunits, two β2-subunits and one γ2-subunit. The general architecture of the receptor is known from structural studies of the pentameric ligand-gated ion channel superfamily[15] and from the structure of a homopentameric $GABA_A$ receptor[16]. In the physiological assembly, GABA binds at β–α-subunit interfaces, and benzodiazepines bind at the α–γ interface[10,17]. Mutagenesis and functional studies have approximated the loci for these and many other compounds on $GABA_A$ receptors[10,17–19], but currently there is no structural information regarding a physiological $GABA_A$ receptor. Here we present high-resolution structures of the α1β2γ2 $GABA_A$ receptor, which illuminate atomic mechanisms of GABA and flumazenil recognition and features of the assembly of this heteromeric receptor.

Biochemistry and structure determination

We optimized receptor constructs and expression conditions to produce and purify the receptor assembly comprising the α1, β2 and γ2-subunits (Methods, Extended Data Fig. 1). We raised monoclonal antibodies to the receptor and purified a receptor–Fab complex to disrupt the low-resolution pseudo-symmetry and facilitate particle alignment[20] (Extended Data Fig. 2a). The purified $GABA_A$ receptor construct (henceforth the 'cryo-EM construct') retained the ability to bind [^{3}H]-flumazenil with low nanomolar affinity[13,21] (Extended Data Fig. 2b). We observed a small positive effect of Fab on GABA potency, and found that binding of Fab did not affect affinity for [^{3}H]-flumazenil. Fab had no effect on the functional response to GABA and flumazenil applied at the concentrations used for cryo-electron microscopy (cryo-EM) (Extended Data Fig. 2).

Processing of cryo-EM images of the sample containing $GABA_A$ receptor, GABA, flumazenil and Fab revealed a homogeneous complex with two bound Fabs (Extended Data Fig. 3). Classification yielded reconstructions with two distinct transmembrane domain (TMD) arrangements, which we call conformation A and conformation B. Refinement of the two reconstructions yielded density maps, both at overall resolutions of approximately 3.9 Å (Extended Data Fig. 4). Cryo-EM density maps were of sufficient quality to allow modelling of almost the entire receptor and the variable domains of the Fabs (Methods and Extended Data Fig. 5–7). The density map shows clear sidechain densities and resolution of 3 Å or better in the extracellular ligand binding sites, whereas the TMD (3–4 Å) and the Fab fragments (4–4.5 Å) are resolved at lower resolutions. The γ2-subunit in conformation B, and in particular its TMD, was comparatively more disordered than the rest of the receptor but still exhibited secondary structural features.

Overall architecture

The $GABA_A$ receptor–Fab complex is a cylinder-shaped receptor assembly, with two Fab fragments extending radially from the extracellular domain (ECD) of the receptor (Fig. 1). Five receptor subunits assemble in a pseudo-symmetrical fashion around an extracellular vestibule and integral ion channel. The two Fab fragments interact

[1]Departments of Neuroscience and Biophysics, University of Texas Southwestern Medical Center, Dallas, TX, USA. *e-mail: ryan.hibbs@utsouthwestern.edu

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

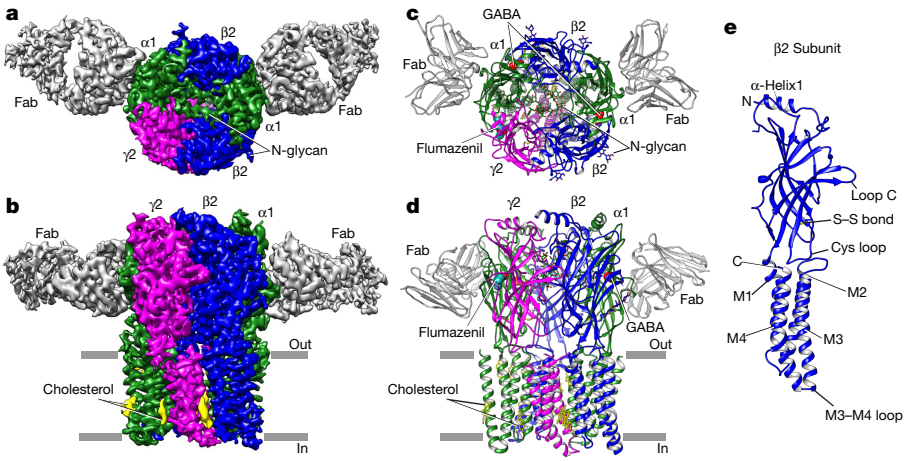


Fig. 1 | Overall structure of the $GABA_A$ receptor–Fab complex. a, **b**, Top and side views of the 3D reconstruction of $GABA_A$ receptor–Fab complex coloured by subunit. α1, green; β2, blue; γ2, magenta; Fab, grey; CHS, yellow. Conformation A is shown. **c**, **d**, Top and side view of the atomic model, coloured as in **a**. Flumazenil is shown as cyan spheres, GABA is shown as red spheres, and CHS and N-linked glycans are modelled as sticks. **e**, Structure of a single β2-subunit.

exclusively with the ECD of the α1-subunits and orient parallel to the membrane (Extended Data Fig. 2h–j). When viewed from the synaptic cleft, the arrangement of subunits around the pentameric ring is β2–α1–β2–α1–γ2 in a counter-clockwise direction, consistent with functional studies of concatameric receptors[22–24]. The ECD of each subunit, which carries the signature Cys-loop, begins with an amino-terminal α-helix at the apex of the pentameric ring, followed by ten β-strands folded into a β-sandwich. The ECD is followed by four α-helices (M1–M4), with M2 lining the ion channel (Fig. 1e).

GABA is bound in the classical neurotransmitter site at each of the two β2–α1 interfaces in the ECD (Fig. 1c). Strong density for flumazenil was observed at an analogous position at the single α1–γ2 interface (Extended Data Figs. 6l, 7l). In addition to the conserved N-linked glycans on the periphery of the β-subunits[16], the extracellular vestibule is populated with a branched network of N-linked glycans emanating from α1-subunits (Fig. 1c). Several ordered densities were observed at the membrane–receptor interface, and at subunit interfaces in the TMD. We have modelled these as CHS, a water-soluble proxy for cholesterol. The ECDs are qualitatively identical between conformations A and B, with the exception of the loops that interact with the TMD (root mean square deviation (r.m.s.d.) of Cα atoms for the entire ECD superposition is 0.34 Å). The TMD conformations are distinct, but both conformations rest in what we suggest are non-conducting desensitized states, based on patch-clamp electrophysiology experiments carried out to achieve steady-state currents (Extended Data Fig. 2e).

Neurotransmitter and benzodiazepine binding sites

We performed patch-clamp experiments to measure the response of the receptor cryo-EM construct to neurotransmitter. We found that application of GABA induced inward currents that were inhibited by bicuculline, a competitive antagonist, in a manner similar to that observed at the wild-type receptor[25,26] (Fig. 2a, Extended Data Fig. 2d).

The $GABA_A$ receptor has two equivalent neurotransmitter-binding pockets located at the β2–α1 interfaces in the ECD (Fig. 2c). We observed strong density in both sites, correlating in size and shape to GABA (Fig. 2b, e, f). The density and chemical environment permitted orientation of the carboxylate and amine ends of the molecule. The principal (+) side of the neurotransmitter pocket contributes mainly aromatic residues: Y97 on loop A, Y157 on loop B, and F200 and Y205 on loop C (Fig. 2e, f). These side chains form an aromatic glove around the basic amino nitrogen of GABA, with F200 and Y205 positioned to make favourable cation–π interactions. Substitutions of Y157 and Y205 greatly decreased the sensitivity to GABA, supporting the importance of these residues for GABA recognition[27]. E155 on loop B is modelled outside of strong interaction distance with the GABA amino nitrogen (4.2 Å), but may further anchor this end of the ligand, consistent with the finding that mutating this residue reduces GABA potency[28]. Finally, T202 is positioned to form a hydrogen bond with the carboxylate group of GABA.

The structure also allowed us to define contributions from the complementary α-subunit. F65 forms the floor of the binding pocket and contributes important hydrophobic interactions, as suggested by

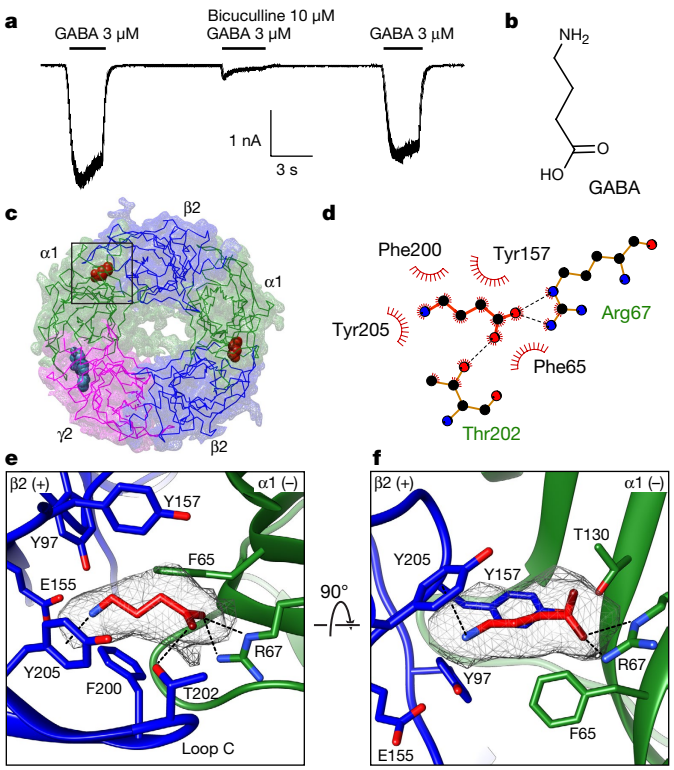


Fig. 2 | Neurotransmitter-binding site. a, Electrophysiology of the $GABA_A$ receptor cryo-EM construct in HEK cells. $n = 3$ independent experiments. **b**, Chemical structure of GABA. **c**, View of the $GABA_A$ receptor construct from the synapse. The box indicates one of two equivalent GABA binding sites at β2–α1 interfaces. **d**, LigPlot schematic of GABA interactions, showing electrostatic (dashes) and hydrophobic (eyelashes) interactions. **e**, **f**, Detailed architecture of the GABA-binding pocket highlighted in **c**. **e**, Synaptic view as in **c**. Putative hydrogen bonding and cation–π-interactions are represented as dashed lines. **f**, Side view of the GABA-binding pocket with loop C backbone hidden for clarity.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

mutagenesis and cysteine crosslinking studies[29,30]. Whereas T130 may contribute a hydrogen bond with the GABA carboxylate, R67 appears to be key to GABA recognition, stabilizing the carboxylate head of the ligand through electrostatic interactions with its basic guanidinium group. Furthermore, this arginine is conserved across α-subunits, but is absent in β-and γ-subunits, and mutating it decreases GABA potency[31]. In β-subunits, the equivalent residue is a glutamine—this difference could account for the weak binding of GABA by GABA–β3 homopentamers[16,32]. Therefore, GABA binding is coordinated through interactions with conserved aromatic residues and electrostatic interactions with side chains complementing the anionic and cationic ends of the neurotransmitter. These interactions suggest that variability on the complementary subunit has a large influence on ligand selectivity.

Benzodiazepines are a class of psychoactive drugs that enhance the effect of GABA at the $GABA_A$ receptor, resulting in sedative, hypnotic, anxiolytic, anticonvulsant and muscle relaxant effects[10]. Classical benzodiazepines potentiate agonist-mediated activation of the $GABA_A$ receptor by causing a decrease in the concentration of GABA required for activation. Flumazenil, an imidazobenzodiazepine (Fig. 3), is a competitive benzodiazepine antagonist. We tested the effect of flumazenil on our cryo-EM constructs and found that it blocks diazepam potentiation of the GABA response and binds with nanomolar affinity (Fig. 3a, Extended Data Fig. 2c).

Including flumazenil in purification of the $GABA_A$ receptor allowed us to interrogate atomic interactions at the benzodiazepine site. At the α1–γ2 interface in the ECD, we observed strong planar density consistent with the geometry of flumazenil (Fig. 3e, f). The shape of the density map accommodated the benzene and imidazole extensions off the diazepine ring when the ligand was positioned in only one orientation. The benzodiazepine ring system sits roughly parallel to loop C, with its fluorobenzene end interacting exclusively with the principal (α1)-subunit and its ethylcarboxylate extending towards the solvent, between the tip of loop C and the complementary (γ2)-subunit. Flumazenil is nestled in an aromatic box formed by three residues from the principal subunit and two from the complementary subunit (Fig. 3d–f). F100 and Y160 from α1 and F77 from γ2 form the back wall of the box; the diazepine ring packs against the phenyl ring of F77. Mutagenesis studies support the aromatic nature of this residue being important for binding: mutation of F77 to tyrosine has little effect on flumazenil affinity[33], but other mutations result in large decreases in affinity[33,34]. Y58 on the complementary subunit packs against the diazepine methyl group and situates on a loop that we designate D′. It forms an incomplete floor of the box; much of the membrane side of the binding pocket is open to solvent, suggesting that the receptor does not require a substantial conformational change to allow flumazenil to dock or undock from its site. Y210 on loop C forms π–π-stacking interactions with the benzene ring of flumazenil, thereby contributing to the front wall of the box. S205, S206 and T207, also on loop C, add to the front wall of the box and may form electrostatic interactions with the flumazenil diazepine, imidazole and ester groups (Fig. 3d, e). T142 in the back of the pocket is also positioned to form a hydrogen bond (2.7 Å) with the flumazenil ester carbonyl oxygen.

Of note, the distal fluorine of the antagonist is positioned to form a hydrogen bond with the H102 indole nitrogen (3.1 Å) from the principal subunit. H102 was identified early as being important in benzodiazepine and flumazenil binding. αβγ receptors that contain α1–α3 or α5 subunits have this histidine and are benzodiazepine-sensitive, whereas α4 and α6-containing receptors have an arginine at this position and do not respond to benzodiazepines[35]. Mutation of this residue to any other tolerated residue decreases the affinity of all benzodiazepines examined as well as flumazenil[35–37], with the exception of histidine to cysteine mutation in the α5-subunit[38]. Analysis of the structure suggests a straightforward explanation for the exquisite sensitivity of flumazenil affinity to the identity of this residue. A residue with a long side chain, such as arginine, would clash with flumazenil in its observed location, and any residue that is unable to form a hydrogen bond with the halide would be unfavourable. Manual superpositioning

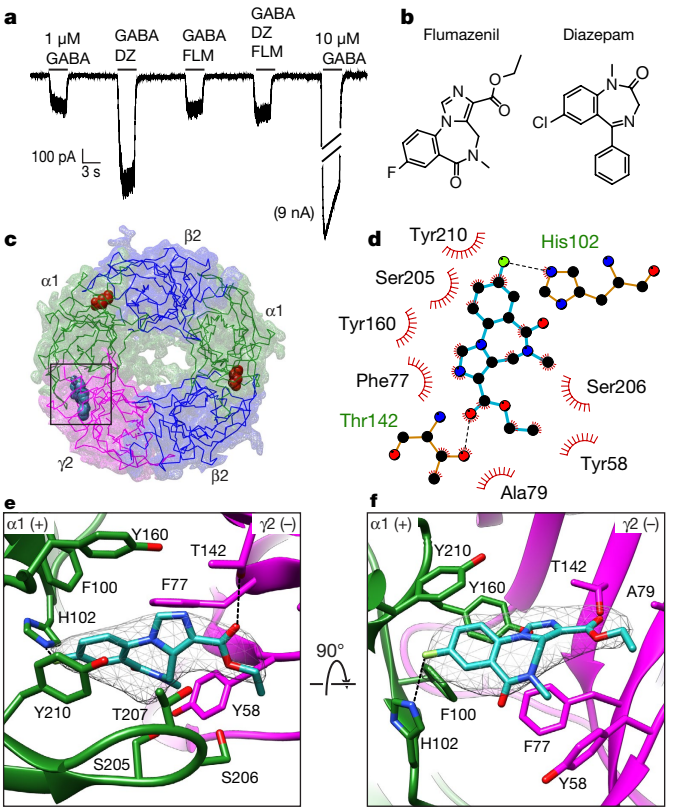


Fig. 3 | Flumazenil interactions at the benzodiazepine binding site. **a**, Electrophysiology of the cryo-EM construct, showing that flumazenil (FLM) (3 μM) blocks $GABA_A$ receptor potentiation by diazepam (DZ) (1 μM). $n = 3$ independent experiments. **b**, Chemical structures of flumazenil and diazepam. **c**, View of the $GABA_A$ receptor from the synapse, as in Fig. 2c. The box indicates flumazenil bound at the α1–γ2 interface. **d**, Schematic of flumazenil interactions showing electrostatic (dashes) and hydrophobic (eyelashes) interactions. **e**, **f**, Detailed architecture of flumazenil-binding pocket highlighted in **c**, with orientations and representations as in Fig. 2e, f.

of the benzodiazepine rings of diazepam on flumazenil suggests that this positive modulator can be accommodated in the same pocket, maintaining the halogen–H102 interaction, without substantial conformational rearrangement (Extended Data Fig. 2f, g). An unanswered question relates to how the antagonist remains inert with respect to GABA binding and activation while the chemically similar benzodiazepines potentiate activation by GABA.

Pseudo-ligand binding sites

There are two ECD interfaces that lack density for ligands in our structure, α1–β2 and γ2–β2 (Fig. 4a). These are potential targets for design of novel modulators but, to our knowledge, only one compound has been characterized as a positive modulator that binds at the α–β-interface[39]. Comparison of the architecture of the empty sites with the occupied GABA and benzodiazepine sites allows for clarification of important binding determinants for these classes of ligands. Superposition of vacant and occupied ligand binding sites reveals an overall conservation of backbone conformation (Fig. 4b–g, Extended Data Fig. 8). All the interface classes contain four aromatic residues at conserved positions: F/Y on loop A, Y on loop B, Y on loop C and F/Y on loop D (Fig. 4h, underlined). Their side chains adopt similar orientations regardless of the presence or absence of ligand. These residues are likely to contribute to the core architecture of $GABA_A$ receptor binding sites, with other residues defining ligand selectivity. Surprisingly, there are no substantial differences in the positions of loop C between the agonist, benzodiazepine, and pseudo-ligand binding pockets (Fig. 4, Extended Data Fig. 8). An illustration of this

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

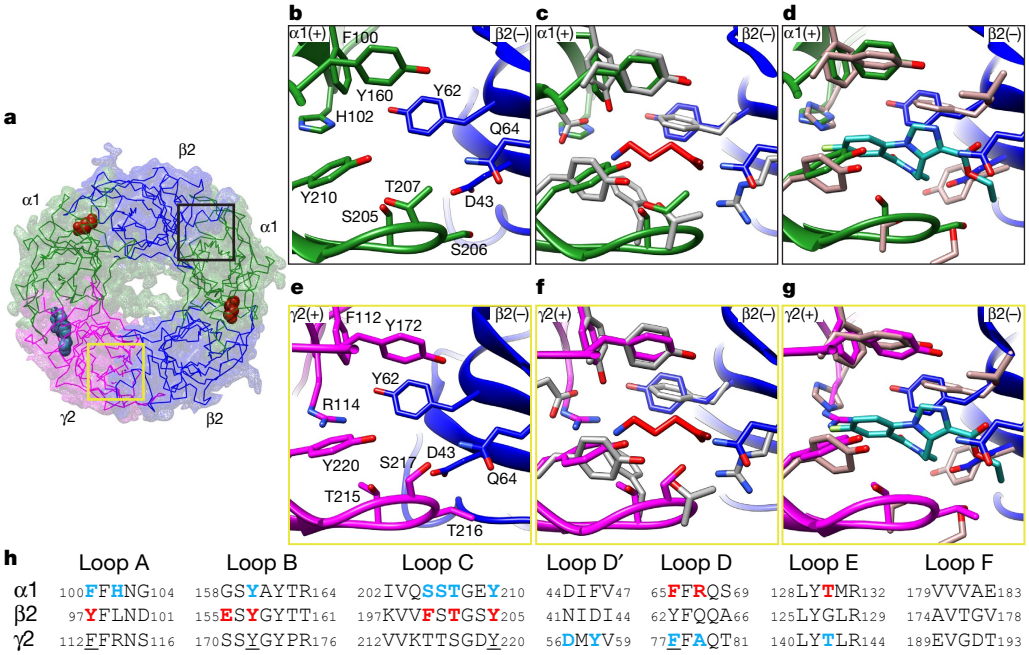


Fig. 4 | Pseudo-ligand sites versus GABA and flumazenil sites. **a**, Synaptic view, with coloured boxes indicating the distinct pseudo-agonist sites. **b**, **e**, Detailed structural information of pseudo-agonist interfaces α–β and γ–β highlighted in **a**. **c**, **f**, Superposition of GABA-binding site (grey) on α–β and γ–β-interfaces, respectively. **d**, **g**, Superposition of flumazenil-binding site (taupe) on the α–β and γ–β-interfaces, respectively. **h**, Sequence alignment of the loops involved in ligand-binding pockets. Blue residues are involved in flumazenil binding, red residues are involved in GABA binding.

ligand-insensitive symmetry is found by comparing the backbone conformations of two α1-subunits, one that forms the empty α–β-interface and one that forms the flumazenil-bound α–γ-interface (Extended Data Fig. 8i, n). Loop C from these subunits may adopt indistinguishable conformations because flumazenil is an inert ligand that simply stabilizes an apo conformation of the α1-subunit, mimicking the empty α–β-interface. Alternatively, conformational changes among subunits in the pentamer may be concerted, with loop C symmetrically 'closed' or 'open'[40].

We used the new structural information to explore the question of why GABA and flumazenil do not bind to the empty α–β and γ–β-interfaces. As described above, the principal aromatics are conserved across all interface classes. The β2-subunit forms the complementary face in both of these pseudo-agonist sites. This subunit lacks the loop D arginine (R67 in α) that appears to be important for high-affinity GABA binding; it is replaced in β2 with a glutamine, which could not form a salt bridge with the GABA carboxylate. This salt bridge may not be a strict requirement, as GABA can bind to β2–β2 interfaces to activate β2γ2 recombinant receptors[41]. F200 on loop C of β2, which contacts the GABA amino nitrogen, is replaced by a polar serine or threonine (α-residue 205, Fig. 4) when α or γ forms the principal face. The lack of π-interactions may destabilize neurotransmitter binding at these interfaces. Moreover, the β2 E155 is replaced with glycine and serine in the α- and γ-subunits, respectively. Thus, both sites lack the charge–charge interactions at both ends of the corresponding GABA position that would promote its binding. Specific to the γ–β-interface, R114 on the γ-subunit would electrostatically repel the GABA amino nitrogen (Fig. 4e, f). This γ2 R114 replaces the α1 H102, akin to the identity at that position in α4 and α6-subunits, which are insensitive to benzodiazepines. In our structure an arginine at this position would clash with flumazenil (Fig. 4g). The methyl side chain of A79 in the γ2-subunit orients towards and accommodates the ethyl carboxylate of flumazenil (Fig. 3f). Its substitution to Q64 in β2 would also clash with flumazenil and is therefore likely to contribute to exclusion of flumazenil from the pseudo-ligand binding interfaces.

Glycosylation and ion permeation

The extracellular vestibule in Cys-loop receptors has been characterized as a wide chamber that is filled with bulk solvent and hydrated ions and has a role in tuning ion conductance[15,42]. All such structures reported to date have shown a cavernous architecture that constricts strongly at the junction of the membrane where it transitions into the ion channel. We observed strong continuous and branched densities extending from consensus glycosylation sites on the α1-subunits inside the vestibule (Fig. 5). Glycan1 comprises eight well-ordered sugars that form an 'H' shape, extending across the channel's central axis to make extensive contacts with the γ2-subunit (Fig. 5b, c). γN101 and γG104 are oriented to form electrostatic interactions with the glycans. The β5 strand (β5–β5′ loop) of the γ2-subunit provides further contacts for anchoring the sugar molecules. Glycan2 adopts a 'Y' shape, consisting of five sugar monomers (Fig. 5d, e). It extends along the β2-subunit (chain C), making contact with glycan1. Mutation of the glycosylation sites results in a loss of expression at physiological temperature[43].

Each glycan chain blocks nearly half of the diameter of the vestibule, leaving a relatively narrow path for ion permeation. The observed gap is approximately 5–8 Å in width, approximating estimates of the open-channel diameter for the Cys-loop receptor superfamily[44,45] (Extended Data Fig. 9). An alternative path for ion entry into the extracellular mouth of the channel is found at subunit interfaces proximal to the membrane. Fenestrations about 7–10 Å wide at a subset of these junctions would allow for hydrated chloride to pass through, as observed in the β3 homopentamer[16] (Extended Data Fig. 9). Thus, permeation through the synaptic $GABA_A$ receptor may involve both the extracellular vestibule and these interfacial fenestrations.

The transmembrane region

We discovered heterogeneity in the cryo-EM data that resulted in two classes of density maps that we call conformations A and B, representing approximately 60% and 40% of our particle dataset, respectively (Extended Data Fig. 3). The major differences between these conformations are restricted to the TMD and to ECD loops that directly contact the TMD (Extended Data Fig. 8). Though they are of great interest, we interpret these conformational differences with caution, as the

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

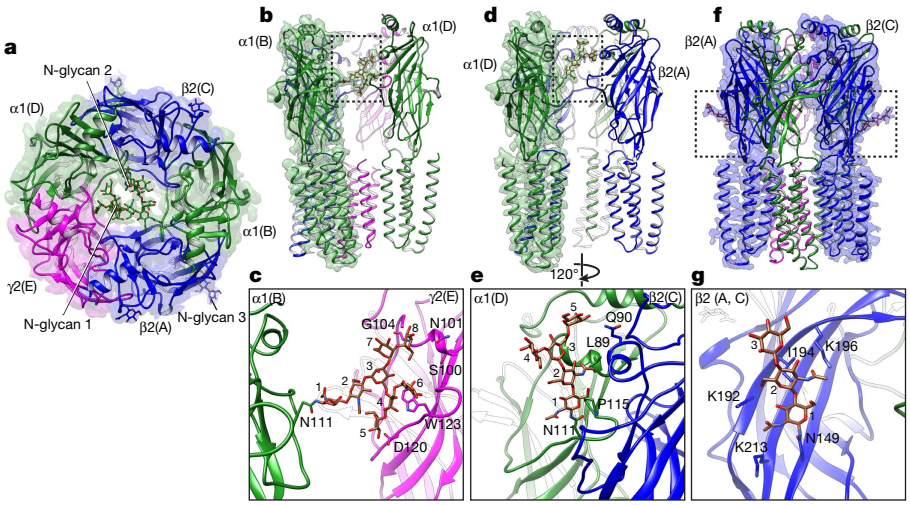


Fig. 5 | Vestibule and non-vestibule N-glycosylation. **a**, Synaptic view of receptor with glycosylation sites indicated. **b**, **d**, Side view of vestibule N-glycosylation attached to α1-subunits. Subunits β2 and γ2 are hidden for clarity. N-linked glycans are indicated by dashed black boxes. **c**, **e**, Detailed structures of N-linked glycans. **c**, NAG_1–NAG_2–BMA_3–MAN_4–MAN_5–MAN_6–MAN_7–MAN_8. **e**, NAG_1–NAG_2–BMA_3–MAN_4–MAN_5. NAG, *N*-acetyl-D-glucosamine; BMA, β-D-mannose; MAN, α-D-mannose. **f**, **g**, Reference and detailed structures of peripheral N-linked glycosylation on β2-subunits. Chain IDs are in parentheses.

structures were obtained in the presence of detergent and the TMD deviates from the five-fold symmetry observed in homopentameric receptors (Fig. 6).

Conformation B is approximately five-fold symmetric in the TMD in comparison to conformation A (Extended Data Fig. 10 a–d). In conformation B, this pseudo-symmetry breaks down locally in the γ2-subunit, which adopts a backbone tertiary fold distinct from that observed in the α1 and β2-subunits (Fig. 6c, Extended Data Fig. 8e). The M3 and M4 helices of the γ2-subunit adopt a unique conformation and tilt about 20–30° relative to other transmembrane helices (Extended Data Fig. 10e). Whereas the density for these two helices is not clear enough to determine the register, we can clearly trace the backbone (Extended Data Fig. 10 a–d). The extracellular end of the M2 helix tilts towards the pore axis, reducing the pore diameter to less than the diameter of a chloride ion (Extended Data Fig. 9). In conformation A, the γ2-subunit collapses into the pore (Fig. 6b), with its M2 helix approximately parallel to the channel axis and blocking the entire length of the transmembrane permeation pathway (Extended Data Fig. 9). The conformation of the TMD of the γ2-subunit requires further analysis to determine its relevance in the physiological gating cycle of the receptor.

Previous studies indicate that the TMD of $GABA_A$ receptors contains a number of solvent-accessible cavities that form the binding sites of numerous drugs and compounds, including barbiturates, anesthetics, neurosteroids and ethanol[19,46]. In the density maps of the TMD for both conformations A and B, we identified a number of sausage-shaped densities that we modelled as CHS, which may hydrolyse and remain associated with hydrophobic regions as cholesterol[47]. In our structures, these molecules are located at membrane–subunit interfaces in the TMD or intercalate between two adjacent subunits, forming extensive contacts with the transmembrane helices (Fig. 6b, d). Further, in both conformations A and B, we found that cholesterol occupies an intra-subunit site at the junction of M3 and M4 helices facing the lipid bilayer. This finding is consistent with recent studies that show binding of endogenous cannabinoids[48] and neurosteroids[49,50] in sites that overlap in part with the cholesterols, suggesting that these sites underlie a common mechanism for modulation of the $GABA_A$ receptor family.

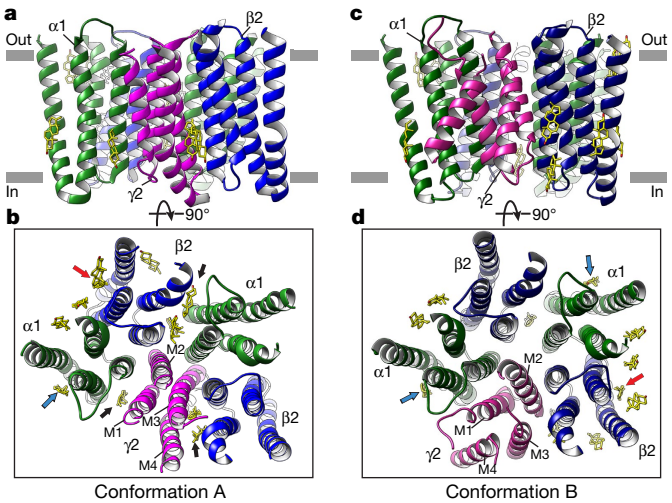


Fig. 6 | Conformational arrangements of transmembrane domain. **a**, **b**, Side and top views of the transmembrane domain in conformation A. **c**, **d**, Side and top views of the transmembrane domain in conformation B. In **b** and **d**, putative cholesterol molecules are shown as sticks. Red arrows indicate sites that overlap with proposed endocannabinoid and pregnenolone sulfate sites. Blue arrows indicate sites shared with pregnenolone sulfate. Black arrows indicate sites shared with the neurosteroids pregnanolone and tetrahydrodeoxycorticosterone (THDOC).

Conclusion

We present high-resolution structures of the predominant synaptic isoform of the human $GABA_A$ receptor. The structures illuminate atomic-scale mechanisms of neurotransmitter binding and GABA selectivity for β–α-subunit interfaces. Analysis of the complex with the benzodiazepine-site antagonist flumazenil reveals principles of selectivity and architecture for this therapeutically important class of drugs. The elucidation of orthosteric and allosteric site recognition provides an expanded blueprint for exploring pharmacologically tractable loci in the other Cys-loop receptor superfamily members. The cryo-EM density maps reveal two distinct transmembrane domain conformations that raise questions for future studies and may relate to the ability of this receptor, more so than other Cys-loop receptor superfamily members, to be modulated by many chemically distinct classes of drugs via its transmembrane domain.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0255-3.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Received: 20 February 2018; Accepted: 17 May 2018;
Published online: 27 June 2018

1. Nutt, D. J. & Malizia, A. L. New insights into the role of the $GABA_A$-benzodiazepine receptor in psychiatric disorder. *Br. J. Psychiatry* **179**, 390–396 (2001).
2. Maconochie, D. J., Zempel, J. M. & Steinbach, J. H. How quickly can $GABA_A$ receptors open? *Neuron* **12**, 61–71 (1994).
3. Braat, S. & Kooy, R. F. The $GABA_A$ receptor as a therapeutic target for neurodevelopmental disorders. *Neuron* **86**, 1119–1130 (2015).
4. Chuang, S. H. & Reddy, D. S. Genetic and molecular regulation of extrasynaptic GABA-A Receptors in the brain: therapeutic insights for epilepsy. *J. Pharmacol. Exp. Ther.* **364**, 180–197 (2018).
5. Jacob, T. C., Moss, S. J. & Jurd, R. $GABA_A$ receptor trafficking and its role in the dynamic modulation of neuronal inhibition. *Nat. Rev. Neurosci.* **9**, 331–343 (2008).
6. Roberts, E. & Frankel, S. γ-Aminobutyric acid in brain: its formation from glutamic acid. *J. Biol. Chem.* **187**, 55–63 (1950).
7. Awapara, J., Landua, A. J., Fuerst, R. & Seale, B. Free γ-aminobutyric acid in brain. *J. Biol. Chem.* **187**, 35–39 (1950).
8. Sternbach, L. H. The benzodiazepine story. *J. Med. Chem.* **22**, 1–7 (1979).
9. Sieghart, W. Allosteric modulation of $GABA_A$ receptors via multiple drug-binding sites. *Adv. Pharmacol.* **72**, 53–96 (2015).
10. Sigel, E. & Buhr, A. The benzodiazepine binding site of $GABA_A$ receptors. *Trends Pharmacol. Sci.* **18**, 425–429 (1997).
11. Votey, S. R., Bosse, G. M., Bayer, M. J. & Hoffman, J. R. Flumazenil: a new benzodiazepine antagonist. *Ann. Emerg. Med.* **20**, 181–188 (1991).
12. Lobo, I. A. & Harris, R. A. $GABA_A$ receptors and alcohol. *Pharmacol. Biochem. Behav.* **90**, 90–94 (2008).
13. Sieghart, W. Structure and pharmacology of γ-aminobutyric $acid_A$ receptor subtypes. *Pharmacol. Rev.* **47**, 181–234 (1995).
14. Sigel, E. & Steinmann, M. E. Structure, function, and modulation of $GABA_A$ receptors. *J. Biol. Chem.* **287**, 40224–40231 (2012).
15. Nemecz, A., Prevost, M. S., Menny, A. & Corringer, P. J. emerging molecular mechanisms of signal transduction in pentameric ligand-gated ion channels. *Neuron* **90**, 452–470 (2016).
16. Miller, P. S. & Aricescu, A. R. Crystal structure of a human $GABA_A$ receptor. *Nature* **512**, 270–275 (2014).
17. Whiting, P. J., McKernan, R. M. & Wafford, K. A. Structure and pharmacology of vertebrate $GABA_A$ receptor subtypes. *Int. Rev. Neurobiol.* **38**, 95–138 (1995).
18. Olsen, R. W. Allosteric ligands and their binding sites define γ-aminobutyric acid (GABA) type A receptor subtypes. *Adv. Pharmacol.* **73**, 167–202 (2015).
19. Forman, S. A. & Miller, K. W. Anesthetic sites and allosteric mechanisms of action on Cys-loop ligand-gated ion channels. *Can. J. Anaesth.* **58**, 191–205 (2011).
20. Walsh, R. M. J. et al. Structural principles of distinct assemblies of the human $\alpha4\beta2$ nicotinic receptor. *Nature* **557**, 261–265 (2018).
21. Marangos, P. J., Patel, J., Boulenger, J. P. & Clark-Rosenberg, R. Characterization of peripheral-type benzodiazepine binding sites in brain using [^{3}H]Ro 5-4864. *Mol. Pharmacol.* **22**, 26–32 (1982).
22. Baumann, S. W., Baur, R. & Sigel, E. Forced subunit assembly in $\alpha1\beta2g\gamma2$ $GABA_A$ receptors. Insight into the absolute arrangement. *J. Biol. Chem.* **277**, 46020–46025 (2002).
23. Baur, R., Minier, F. & Sigel, E. A. $GABA_A$ receptor of defined subunit composition and positioning: concatenation of five subunits. *FEBS Lett.* **580**, 1616–1620 (2006).
24. Tretter, V., Ehya, N., Fuchs, K. & Sieghart, W. Stoichiometry and assembly of a recombinant $GABA_A$ receptor subtype. *J. Neurosci.* **17**, 2728–2737 (1997).
25. Curtis, D. R., Duggan, A. W., Felix, D. & Johnston, G. A. GABA, bicuculline and central inhibition. *Nature* **226**, 1222–1224 (1970).
26. Ueno, S., Bracamontes, J., Zorumski, C., Weiss, D. S. & Steinbach, J. H. Bicuculline and gabazine are allosteric inhibitors of channel opening of the $GABA_A$ receptor. *J. Neurosci.* **17**, 625–634 (1997).
27. Amin, J. & Weiss, D. S. $GABA_A$ receptor needs two homologous domains of the β-subunit for activation by GABA but not by pentobarbital. *Nature* **366**, 565–569 (1993).
28. Mortensen, M. et al. Photo-antagonism of the $GABA_A$ receptor. *Nat. Commun.* **5**, 4454 (2014).
29. Sigel, E., Baur, R., Kellenberger, S. & Malherbe, P. Point mutations affecting antagonist affinity and agonist dependent gating of $GABA_A$ receptor channels. *EMBO J.* **11**, 2017–2023 (1992).
30. Smith, G. B. & Olsen, R. W. Identification of a [^{3}H]muscimol photoaffinity substrate in the bovine γ-aminobutyric $acid_A$ receptor α subunit. *J. Biol. Chem.* **269**, 20380–20387 (1994).
31. Boileau, A. J., Evers, A. R., Davis, A. F. & Czajkowski, C. Mapping the agonist binding site of the $GABA_A$ receptor: evidence for a β-strand. *J. Neurosci.* **19**, 4847–4854 (1999).
32. Goldschen-Ohm, M. P., Wagner, D. A. & Jones, M. V. Three arginines in the $GABA_A$ receptor binding pocket have distinct roles in the formation and stability of agonist- versus antagonist-bound complexes. *Mol. Pharmacol.* **80**, 647–656 (2011).
33. Buhr, A., Baur, R. & Sigel, E. Subtle changes in residue 77 of the γ subunit of $\alpha1\beta2\gamma2$ $GABA_A$ receptors drastically alter the affinity for ligands of the benzodiazepine binding site. *J. Biol. Chem.* **272**, 11799–11804 (1997).
34. Wingrove, P. B., Thompson, S. A., Wafford, K. A. & Whiting, P. J. Key amino acids in the γ subunit of the γ-aminobutyric $acid_A$ receptor that determine ligand binding and modulation at the benzodiazepine site. *Mol. Pharmacol.* **52**, 874–881 (1997).
35. Wieland, H. A., Luddens, H. & Seeburg, P. H. A single histidine in $GABA_A$ receptors is essential for benzodiazepine agonist binding. *J. Biol. Chem.* **267**, 1426–1429 (1992).
36. Dunn, S. M., Davies, M., Muntoni, A. L. & Lambert, J. J. Mutagenesis of the rat $\alpha1$ subunit of the γ-aminobutyric $acid_A$ receptor reveals the importance of residue 101 in determining the allosteric effects of benzodiazepine site ligands. *Mol. Pharmacol.* **56**, 768–774 (1999).
37. Davies, M., Bateson, A. N. & Dunn, S. M. Structural requirements for ligand interactions at the benzodiazepine recognition site of the $GABA_A$ receptor. *J. Neurochem.* **70**, 2188–2194 (1998).
38. Baur, R. & Sigel, E. Replacement of histidine in position 105 in the α_5 subunit by cysteine stimulates zolpidem sensitivity of $\alpha_5\beta_2\gamma_2$ $GABA_A$ receptors. *J. Neurochem.* **103**, 2556–2564 (2007).
39. Ramerstorfer, J. et al. The $GABA_A$ receptor $\alpha+\beta-$ interface: a novel target for subtype selective drugs. *J. Neurosci.* **31**, 870–877 (2011).
40. Unwin, N. Nicotinic acetylcholine receptor and the structural basis of neuromuscular transmission: insights from *Torpedo* postsynaptic membranes. *Q. Rev. Biophys.* **46**, 283–322 (2013).
41. Wongsamitkul, N. et al. alpha subunits in $GABA_A$ receptors are dispensable for GABA and diazepam action. *Sci. Rep.* **7**, 15498 (2017).
42. Hansen, S. B., Wang, H. L., Taylor, P. & Sine, S. M. An ion selectivity filter in the extracellular domain of Cys-loop receptors reveals determinants for ion conductance. *J. Biol. Chem.* **283**, 36066–36070 (2008).
43. Buller, A. L., Hastings, G. A., Kirkness, E. F. & Fraser, C. M. Site-directed mutagenesis of N-linked glycosylation sites on the gamma-aminobutyric acid type A receptor alpha 1 subunit. *Mol. Pharmacol.* **46**, 858–865 (1994).
44. Dwyer, T. M., Adams, D. J. & Hille, B. The permeability of the endplate channel to organic cations in frog muscle. *J. Gen. Physiol.* **75**, 469–492 (1980).
45. Jensen, M. L., Schousboe, A. & Ahring, P. K. Charge selectivity of the Cys-loop family of ligand-gated ion channels. *J. Neurochem.* **92**, 217–225 (2005).
46. Johnston, G. A. $GABA_A$ receptor pharmacology. *Pharmacol. Ther.* **69**, 173–198 (1996).
47. Penmatsa, A., Wang, K. H. & Gouaux, E. X-ray structure of dopamine transporter elucidates antidepressant mechanism. *Nature* **503**, 85–90 (2013).
48. Baur, R. et al. Molecular analysis of the site for 2-arachidonylglycerol (2-AG) on the β_2 subunit of $GABA_A$ receptors. *J. Neurochem.* **126**, 29–36 (2013).
49. Miller, P. S. et al. Structural basis for $GABA_A$ receptor potentiation by neurosteroids. *Nat. Struct. Mol. Biol.* **24**, 986–992 (2017).
50. Laverty, D. et al. Crystal structures of a $GABA_A$-receptor chimera reveal new endogenous neurosteroid-binding sites. *Nat. Struct. Mol. Biol.* **24**, 977–985 (2017).

Acknowledgements We thank D. Cawley for antibody production, X. Bai for electron microscopy discussion, P. Emsley for guidance on glycosylation tools in Coot, W. Costello for initial construct screening and all members of the Hibbs laboratory for discussion. Single-particle cryo-EM data were collected at the University of Texas Southwestern Medical Center Cryo-Electron Microscopy Facility, which is supported by the CPRIT Core Facility Support Award RP170644. We thank D. Nicastro and D. Stoddard for support in facility access and data acquisition. R.W. acknowledges support from the Sara and Frank McKnight Fund for Biochemical Research and the NIH (T32GM008203). R.E.H. is supported by a McKnight Scholar Award, The Welch Foundation (I-1812) and the NIH (DA037492, DA042072, and NS095899).

Reviewer information *Nature* thanks G. Akk, P. Corringer, A. Evers and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions S.Z. performed sample preparation, data collection, model building and writing of the manuscript. C.M.N. and R.M.W. collected microscopy data and edited the manuscript. J.T. performed electrophysiology experiments. J.J.K. assisted in biochemistry and edited the manuscript. R.E.H. oversaw all aspects of the project.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0255-3.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0255-3.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to R.E.H.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Protein construct optimization, expression and purification. The human α1, β2 and γ2 $GABA_A$ receptor subunit genes were codon optimized, synthesized, and cloned into the pEZT-BM expression vector[51]. Enhanced green fluorescent protein (EGFP) was inserted into the M3–M4 loop of each subunit for small scale optimization of constructs and expression conditions. Co-transfections of GnTI$^-$ HEK293S cells (ATCC CRL-3022, not authenticated, not tested for mycoplasma) with combinations of EGFP-tagged and non-tagged subunits were analysed by fluorescence-detection size-exclusion chromatography (FSEC)[52], revealing that whereas α1 and β2-subunits assemble together robustly as heteropentamers, γ2-subunits were only expressed as a part of pentamers including both α1 and β2-subunits. Earlier experiments suggested that when these three subunits are present in a pentamer, the stoichiometry of $\alpha1_2\beta2_2\gamma2_1$ is invariable[22,23]. Thus, a twin strep tag was placed at the N terminus of the γ2-subunit following the predicted signal peptide cleavage site[53]. Several constructs were tested with cryo-EM. The best-ordered density maps resulted from production of receptors wherein a seven amino acid linker[16,79] was used in place of the intracellular loop between transmembrane helices 3 and 4 for all subunits (Extended Data Fig. 1). Residue numbering in the text and atomic models is for the mature (signal-peptide cleaved) human isoforms of all subunits.

Bacmam virus was produced for each subunit and titered as described for the α4β2 nicotinic receptor[51]. A 4–6 l suspension of HEK293S GnTI$^-$ cells[54] was transduced with multiplicities of infection of 0.5:0.5:0.25 for the α1:β2:γ2-subunits, respectively. At the time of transduction, valproic acid (sodium salt, Sigma-Aldrich) was added to 3 mM and flumazenil (Santa Cruz Biotechnology) to 10 μM to boost expression. Protein expression was carried out at 30 °C with 8% CO_2. Cells were collected after ~72 h by centrifugation, resuspended in 20 mM Tris, pH 7.4, 150 mM NaCl (TBS buffer), 1 μM flumazenil, 2 mM γ-aminobutyric acid (GABA; Sigma-Aldrich) and 1 mM phenylmethanesulfonyl fluoride (Sigma-Aldrich), and disrupted using an Avestin Emulsiflex. Lysed cells were centrifuged for 15 min at 10,000*g*; supernatants containing membranes were centrifuged for 2 h at 186,000*g*. Membrane pellets were mechanically homogenized and solubilized for 1 h at 4 °C in a solution containing TBS, 40 mM *n*-dodecyl-β-D-maltopyranoside (DDM; Anatrace), 1 μM flumazenil, and 2 mM GABA. Solubilized membranes were centrifuged for 40 min at 186,000*g*, then passed over Strep-Tactin XT Superflow (IBA) affinity resin. The resin was washed with size exclusion chromatography (SEC) buffer containing TBS, 1 mM DDM, 1 μM flumazenil, 2 mM GABA and 0.2 mM CHS (Anatrace), and eluted in the same buffer containing 50 mM biotin (Sigma-Aldrich).

Generation of monoclonal antibodies and Fab fragments. The 1F4 monoclonal antibody (mAb) against the α1β2γ2 $GABA_A$ receptor (IgG2b, κ) was raised using standard methods (Monoclonal Core, Vaccine and Gene Therapy Institute, Oregon Health & Science University). High affinity and specificity of the antibody for properly folded pentameric $GABA_A$ receptor was assayed by FSEC with EGFP-tagged receptor (shift in elution volume) and western blot (no binding). Fab fragments were generated by papain cleavage of whole antibody at a final concentration of 1 mg/ml for 2 h at 37 °C in 50 mM $NaPO_4$, pH 7.0, 1 mM EDTA, 10 mM cysteine and 1:30 (w/w) papain. Digestion was quenched using 30 mM iodoacetamide at 25 °C for 10 min. Fab was purified by anion exchange using a HiTrap Q HP (GE Healthcare) column in 10 mM Tris, pH 8.0 and a NaCl gradient elution. Cloning and sequencing of Fab antibody regions were performed from mouse hybridoma cells.

Cryo-EM sample preparation. Purified $GABA_A$ receptor from affinity chromatography was mixed with Fab in a 3:1 (w/w) ratio and injected over a Superose 6 Increase 10/300 GL column (GE Healthcare) equilibrated in SEC buffer. Peak fractions were assayed by SEC, monitoring tryptophan fluorescence. The peak fraction was concentrated tenfold to 5–6 mg/ml. Three microlitres of purified $GABA_A$ receptor + GABA + flumazenil + Fab complex was applied to glow-discharged gold R1.2/1.3 200 mesh holey carbon grids (Quantifoil) and immediately blotted for 3 s at 100% humidity and 4 °C, then plunge-frozen into liquid ethane cooled by liquid nitrogen using a Vitrobot Mark IV (FEI).

Cryo-EM image collection and processing. Electron microscopy images were collected using a Titan Krios electron microscope (FEI) operated at an acceleration voltage of 300 kV. Dose-fractionated images were recorded on a K2 Summit direct electron detector (Gatan) equipped with GIF quantum energy filter (20 eV) (Gatan) in super-resolution mode (super-resolution pixel size: 0.535 Å/pixel). 5,594 images were collected over two 72-h sessions. Each micrograph was exposed for 10 s with a dose rate of 4.7 e^- $Å^{-2}$ sec^{-1} (total specimen micrograph dose, 47 e^- $Å^{-2}$), with 40 frames per micrograph. Images were recorded using the automated-acquisition program EPU (FEI) with set defocus values ranging from −1.8 μm to −3 μm.

Dose-fractionated images (movies) were gain normalized, 2× Fourier binned (resulting in a pixel size of 1.07), aligned, dose-weighted and summed using MotionCor2[55]. Defocus values were estimated using GCTF[56]. A total of 1,050,737 auto-picked particles were subjected to 2D classification to remove false-positive and defective particles. The particle set then underwent 3D classification, resulting in 494,727 particles for particle polishing in Relion[57]. 2D classification was used after particle polishing to obtain a final set of 493,104 polished particles.

3D refinement of this polished particle set resulted in a map with clear backbone and side chain density for all of the ECD, Fab variable fragments and the TMDs of the α-subunits. There was obvious structural heterogeneity or disorder in the TMDs of the β- and γ-subunits. We successfully resolved this heterogeneity using the output orientations from the 3D refinement of the polished particle set to perform 3D classification with local angular sampling[58]. This local angular sampling approach revealed improved TMD features with particles segregating into two distinct TMD conformations. 292,662 polished particles from five 3D classes were pooled for 3D refinement in Relion and yielded a reconstruction of the $GABA_A$ receptor conformation A at 3.92 Å overall resolution (Fourier shell correlation (FSC) = 0.143). The remaining 200,442 particles from three 3D classes were combined and yielded a reconstruction of the $GABA_A$ receptor conformation B also at 3.86 Å overall resolution. ResMap[59] was used to estimate local resolution.

Model building, refinement and validation. The nominally higher-resolution model of conformation B was built first, focusing initially on the well-ordered ECD. A homology model was made for each subunit using the crystal structure of the β3 homopentameric $GABA_A$ receptor (PDB ID: 4COF) via Swiss-Model[60]. From side chain and glycosylation features in the ECDs we were able to unambiguously assign the α1, β2 and γ2-subunits; rigid docking of the homology models into these densities supported the assignments. Chain IDs in the models are: A, β2; B, α1; C, β2; D, α1; E, γ2. Density for Fab fragments was observed extending from the ECD of α1-subunits roughly parallel to the membrane plane. Swiss-Model was used to generate a homology model of the Fab light chain using PDB entry 1UYW and of the heavy chain using PDB entry 4WEB and these chains were docked into the cryo-EM density at one Fab site using Chimera[61]. Manual adjustments of the receptor-Fab structure were then performed in Coot[62,63]. The ECD and TMD halves of each subunit and the variable half of each Fab were rigid body fitted into the density map. The variable domain of the Fab was rebuilt into unambiguous density; the density associated with the Fab constant domain was too disordered for both Fab copies to allow building of an atomic model; as this portion of the Fab was not of biological interest, the final model includes just the Fab variable domains. Once this first Fab copy was rebuilt, it was copied into the additional site on the second α1-subunit and manually adjusted. Well-ordered N-linked glycans were built in the ECD channel vestibule and shorter N-linked chains along the outer surface of the ECD. Strong density for ligands was observed at β–α interfaces (modelled as GABA) and at the α–γ interface (flumazenil). The extended conformation of flumazenil is consistent with its crystal structure[64]. No unaccounted-for density was observed at the α–β or γ–β interfaces in either conformation. The ECD of conformation B was docked into the map for conformation A; we noted no meaningful conformational differences except in the loops contacting the TMD.

In conformation A, the TMD portions of the α1-subunits (chains B, D) are well ordered and there was no ambiguity regarding register. The β2-subunits are less well-ordered but we are still confident about register as the backbone adopts internally consistent conformations. In the γ-subunit, which undergoes a conformational rearrangement to fill the pore with its M2 helix, the linker connecting β10 to M1 and the M2–M3 loop are disordered. Nonetheless, the four-helix bundle from this subunit holds together in a conformation akin to the well-ordered subunits, and to that observed in the β3 homopentameric structure, and thus we are confident about the amino acid register in the γ-subunit TMD in conformation A.

In conformation B, the TMD is comparatively less well ordered than in conformation A. The α1-subunits remain, as in conformation A, well ordered, with clear side chain density. The β2-subunit at the chain A position is well ordered, however the chain C β2-subunit, which packs opposite the pore from the γ2-subunit, is not well ordered. Nonetheless, its conformation in less sharpened maps was clear enough to dock the four-helix bundle in a conformation similar to that observed in chain A. The γ2-subunit in conformation B is comparatively disordered and its modelling is problematic as the helix bundle is not held together in a familiar arrangement. We modelled the M1 and M2 helices with amino acids placed tentatively based on side chain density. The M3 and M4 helices were built as poly-alanine chains. Due to a lack of strong interaction of the γ2-subunit in this conformation with its ECD half, the tip of the γ2-subunit Cys-loop did not have clear density and we omitted three residues from this region. Otherwise, the ECD modelling is continuous from the first amino acid to the end of β10. In the TMD, we modelled many strong oblong features as CHS; these occupy distinct sites between the two receptor conformations. After manual building in Coot, global real space coordinate and B-factor refinement with NCS restraints were performed in Phenix[65]. The refined model quality was assessed using Molprobity (Extended Data Fig. 5). The following segments of the receptor were not modelled due to weak density features in the corresponding regions: in conformation A, β2 (chain A and C): N–7, 341–Cα1 (chain B): N–9, 346–C; α1 (chain D): N–9, 348–C; γ2

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

(chain E): N–24, 233–236, 287–291, 356–C; in conformation B, β2 (chain A and C): N–7, 341–C; (chain B): N–12, 346–C; α1 (chain D): N–10, 348–C; γ2 (chain E): N–24, 158–160, 288–296, 319–326, 347–C.

The validation to test for overfitting of the model was performed as previously described[66]. Briefly, the atom positions of the final refined models were randomly displaced by a maximum of 0.5 Å using PDBSET in the CCP4 suite[67]. This perturbed model was then refined in Phenix in real space against the first half map of the reconstruction comprising 50% of the particles. A map versus model FSC comparison was made for this model versus the map used in its refinement ('work'), as well as the same model versus the half map not used in refinement ('free'). The FSC curves of work and free half maps versus model agree well (Extended Data Fig. 5).

Schematic interaction analysis of GABA and flumazenil was performed by Ligplot+[68]. Subunit interfaces were analysed using the PDBePISA server[69]. Structural biology software packages were compiled by SBGrid[70].

For display settings in figures, density maps for the ECD were sharpened as shown in Extended Data Fig. 5a with a B factor −186 Å^2 for conformation A and −153 Å^2 for conformation B. Density maps displayed for TMDs were sharpened with a B factor of −100 Å^2 for both conformations. In Figs. 2, 3, density maps for GABA and flumazenil were rendered in Chimera at threshold levels of 0.018 and 0.04, respectively. In Extended Data Fig. 6, density maps were displayed at the following threshold levels: a–h, 0.024; i–k, 0.0158; l, 0.06; m and n, 0.03. In Extended Data Fig. 7, density maps were displayed at the following threshold levels: a–h, 0.024; i–k, 0.015; l, 0.05; m and n, 0.027. In Figs. 1, 5, Extended Data Fig. 9, density maps were rendered in Chimera at a threshold level of 0.024. In Extended Data Fig. 10a–d, density maps were rendered in Chimera at a threshold level of 0.02.

Electrophysiology. Whole-cell voltage-clamp recordings were made from cells transiently transfected with the constructs used in structural analysis. For the patch-clamp experiments, adherent HEK293S GnTI$^-$ cells were transiently transfected with pEZT-based plasmids 2–3 days before recording. Each 35-mm dish of cells was transfected with the DNA of the α1-subunit, β2-subunit and γ2-subunit in a ratio of 1:1:4 to ensure the incorporation of γ2-subunit. At the time of transfection, cells were moved to 30 °C. On the day of recording, cells were washed with bath solution, which contains (in mM): 140 NaCl, 2.4 KCl, 4 $MgCl_2$, 4 $CaCl_2$, 10 HEPES pH 7.3 and 10 glucose. Borosilicate pipettes were pulled and polished to a resistance of 2–4 MΩ. The pipette solution contained (in mM): 150 CsCl, 10 NaCl, 10 EGTA, 20 HEPES pH 7.3. Cells were clamped at –75 mV. The recordings were made with an Axopatch 200B amplifier, low pass filtered at 2 kHz and digitized at 5 kHz using the Digidata 1440A and pClamp 10 software (Molecular Devices). The GABA, flumazenil, diazepam (Tocris Bioscience) and bicuculline (Sigma-Aldrich) solutions were prepared in bath solution. Stock solution of 1 M GABA was prepared in water; stock solutions of 100 mM bicuculline, 10 mM diazepam and 10 mM flumazenil were prepared in DMSO. Solution exchange was achieved using a gravity driven RSC-200 rapid solution changer (Bio-Logic).

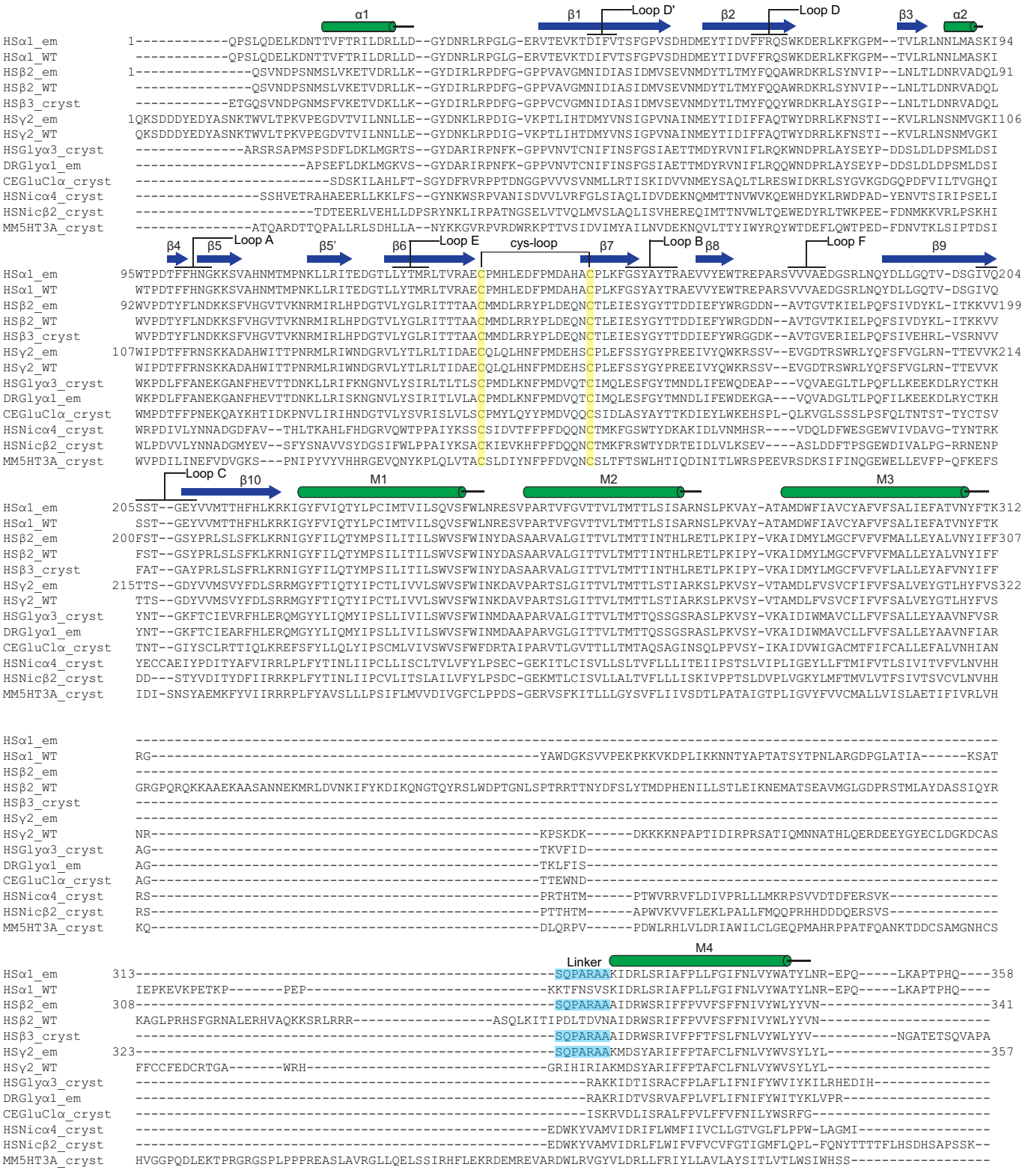
Radioligand binding. Experiments to measure binding of [^{3}H]-flumazenil (PerkinElmer, 84.4 Ci/mmol) to the $GABA_A$ receptor were performed with protein purified as for cryo-EM but in the absence of flumazenil, GABA or Fab. The concentration of binding sites was kept at 1 nM after a series of preliminary experiments to test the optimal protein concentration. For the binding experiments with +Fab sample, Fab was added in excess (100 nM). In addition to the receptor, the binding assay conditions included 20 mM Tris pH 7.4, 150 mM NaCl, 1 mM DDM, and 1 mg/ml streptavidin-YiSi scintillation proximity assay beads (SPA; GE Healthcare Life Sciences). Non-specific signal was determined in the presence of 100 μM [^{1}H]-flumazenil. All data shown are from background-subtracted measurements. For radioligand competition experiments to measure the K_i of diazepam, binding site concentration was 10 nM and the concentration of [^{3}H]-flumazenil was also 10 nM. Each set of binding reaction experiments was performed 2–3 times, in triplicate. The data were analysed using Prism 6 software (GraphPad).

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. Atomic coordinates of the two $GABA_A$ receptor–GABA–flumazenil–Fab complexes have been deposited in the Protein Data Bank with accession number 6D6U for conformation A and 6D6T for conformation B. The cryo-EM density maps have been deposited in the Electron Microscopy Data Bank with accession number EMD-7817 for conformation A and EMD-7816 for conformation B.

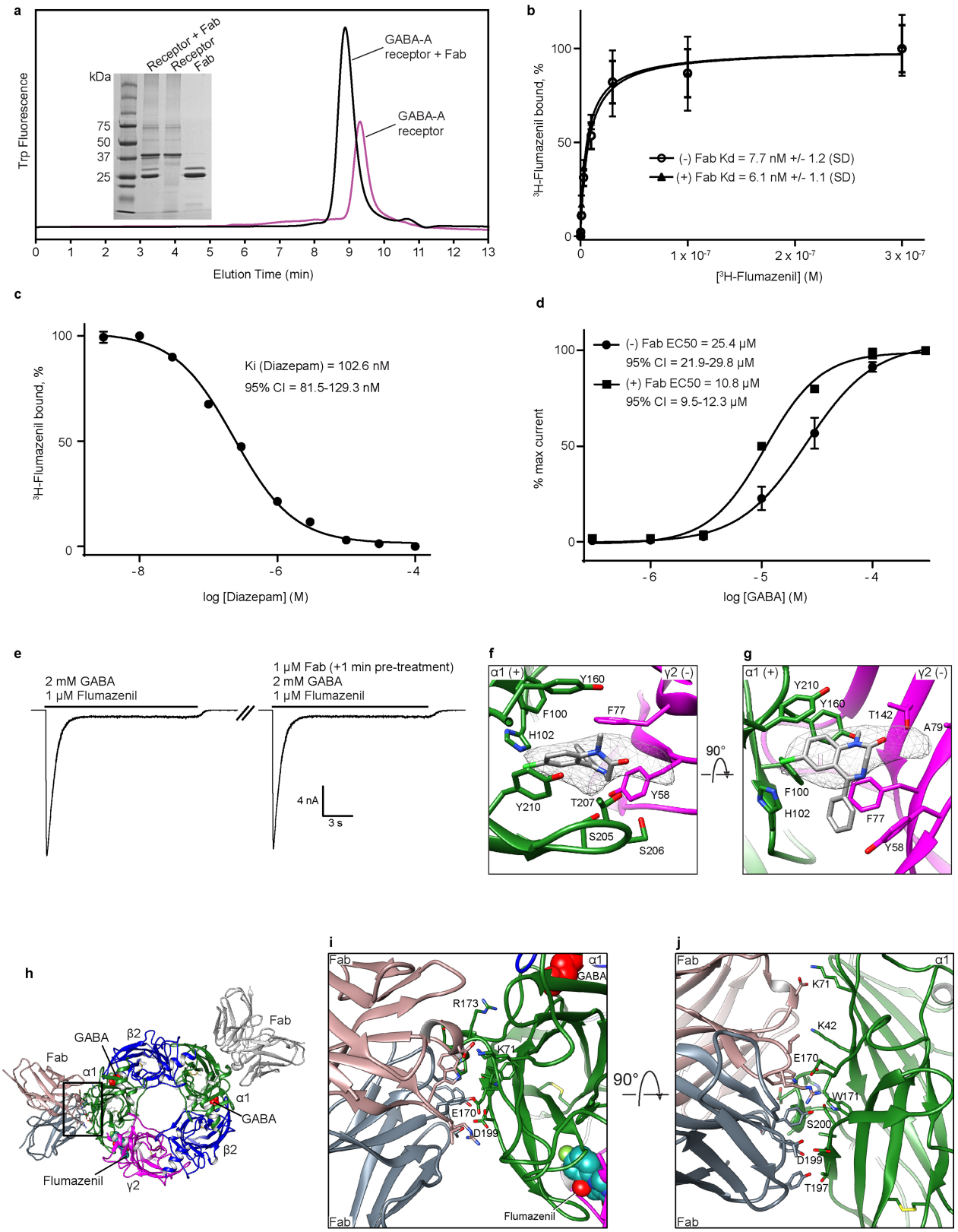
51. Morales-Perez, C. L., Noviello, C. M. & Hibbs, R. E. Manipulation of subunit stoichiometry in heteromeric membrane proteins. *Structure* **24**, 797–805 (2016).
52. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
53. Nielsen, H. Predicting secretory proteins with SignalP. *Methods Mol. Biol.* **1611**, 59–73 (2017).
54. Reeves, P. J., Callewaert, N., Contreras, R. & Khorana, H. G. Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible *N*-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc. Natl Acad. Sci. USA* **99**, 13419–13424 (2002).
55. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
56. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
57. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
58. Scheres, S. H. Processing of structurally heterogeneous cryo-EM data in RELION. *Methods Enzymol.* **579**, 125–157 (2016).
59. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
60. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **31**, 3381–3385 (2003).
61. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
62. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
63. Emsley, P. & Crispin, M. Structural analysis of glycoproteins: building N-linked glycans with Coot. *Acta Crystallogr D Struct Biol* **74**, 256–263 (2018).
64. Codding, P. W. & Muir, A. K. Molecular structure of Ro15-1788 and a model for the binding of benzodiazepine receptor ligands. Structural identification of common features in antagonists. *Mol. Pharmacol.* **28**, 178–184 (1985).
65. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
66. Amunts, A. et al. Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014).
67. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
68. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
69. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
70. Morin, A. et al. Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).
71. Hadingham, K. L. et al. Role of the beta subunit in determining the pharmacology of human gamma-aminobutyric acid type A receptors. *Mol. Pharmacol.* **44**, 1211–1218 (1993).
72. Karim, N. et al. Potency of GABA at human recombinant $GABA_A$ receptors expressed in *Xenopus* oocytes: a mini review. *Amino Acids* **44**, 1139–1149 (2013).
73. Mortensen, M., Patel, B. & Smart, T. G. GABA potency at $GABA_A$ receptors found in synaptic and extrasynaptic zones. *Front. Cell. Neurosci.* **6**, 1 (2011).
74. Baur, R. & Sigel, E. On high- and low-affinity agonist sites in $GABA_A$ receptors. *J. Neurochem.* **87**, 325–332 (2003).
75. Richter, L. et al. Diazepam-bound $GABA_A$ receptor models identify new benzodiazepine binding-site ligands. *Nat. Chem. Biol.* **8**, 455–464 (2012).
76. Middendorp, S. J., Puthenkalam, R., Baur, R., Ernst, M. & Sigel, E. Accelerated discovery of novel benzodiazepine ligands by experiment-guided virtual screening. *ACS Chem. Biol.* **9**, 1854–1859 (2014).
77. Du, J., Lu, W., Wu, S., Cheng, Y. & Gouaux, E. Glycine receptor mechanism elucidated by electron cryo-microscopy. *Nature* **526**, 224–229 (2015).
78. Hibbs, R. E. & Gouaux, E. Principles of activation and permeation in an anion-selective Cys-loop receptor. *Nature* **474**, 54–60 (2011).
79. Jansen, M., Bali, M. & Akabas, M. H. Modular design of Cys-loop ligand-gated ion channels: functional 5-HT3 and GABA rho1 receptors lacking the large cytoplasmic M3M4 loop. *J. Gen. Physiol.* **131**, 137–146 (2008).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 1 | Alignment of $GABA_A$ and other Cys-loop receptor subunits. Cryo-EM constructs (γ2 affinity tag not shown) are numbered starting with the first residue of the mature protein. Sequences aligned (UniProt or PDB accession codes): *Homo sapiens* α1 $GABA_A$ (HS, P14867), *H. sapiens* β2 $GABA_A$ (P47870), *H. sapiens* γ2 $GABA_A$ (P18507), *H. sapiens* $GABA_A$ β3 (4COF), *H. sapiens* glycine α3 (5CFB), *Danio rerio* glycine α1 (DR, 3JAE), *Caenorhabditis elegans* α (CE, 3RHW), *H. sapiens* α4 nAChR (5KXI), *H. sapiens* β2 nAChR (5KXI) and *Mus musculus* 5-HT_3 receptor (MM, 4PIR). α-helices (cylinders), β-strands (arrows), and inserted linker (cyan) are indicated.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

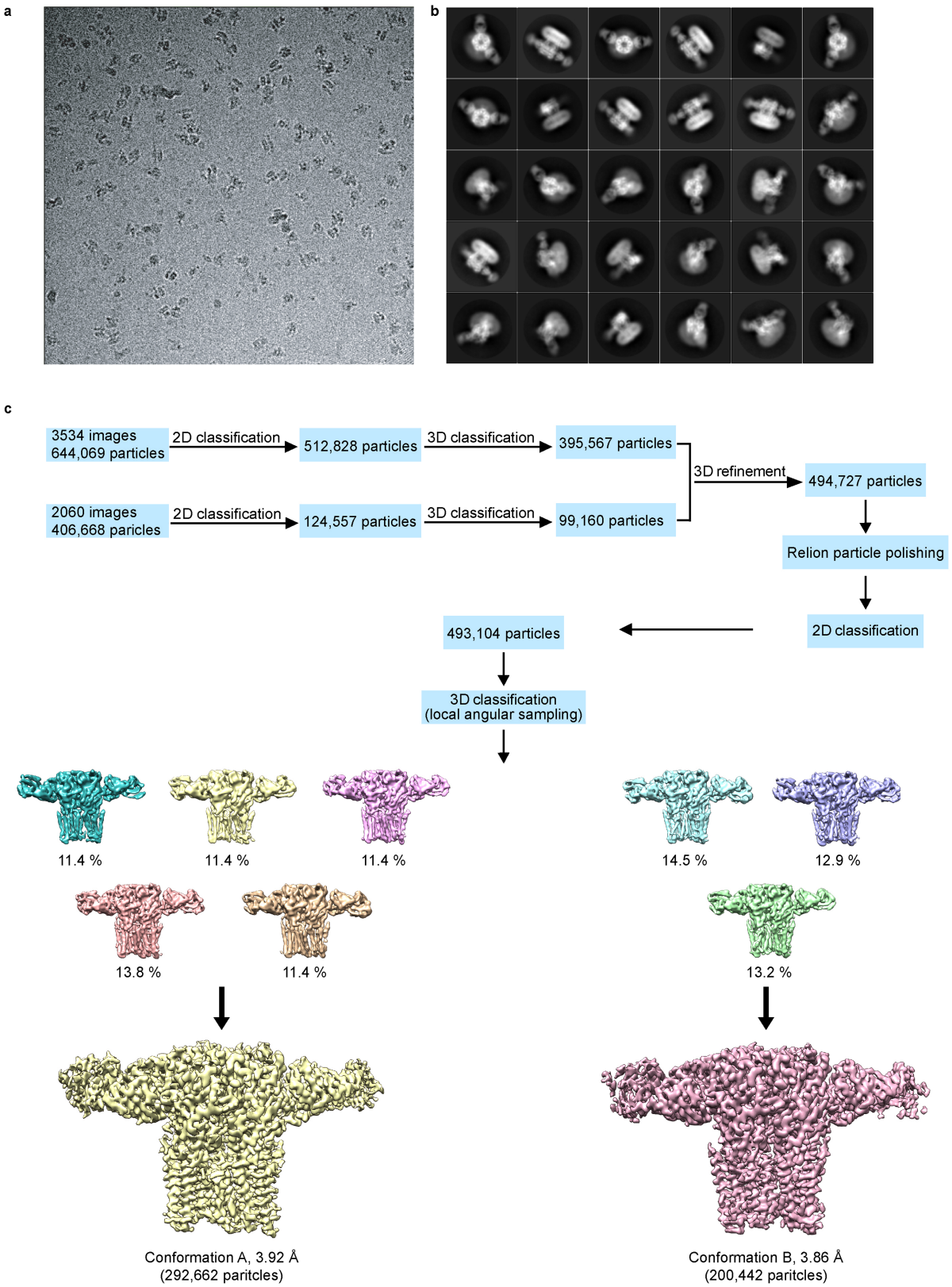


Extended Data Fig. 2 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

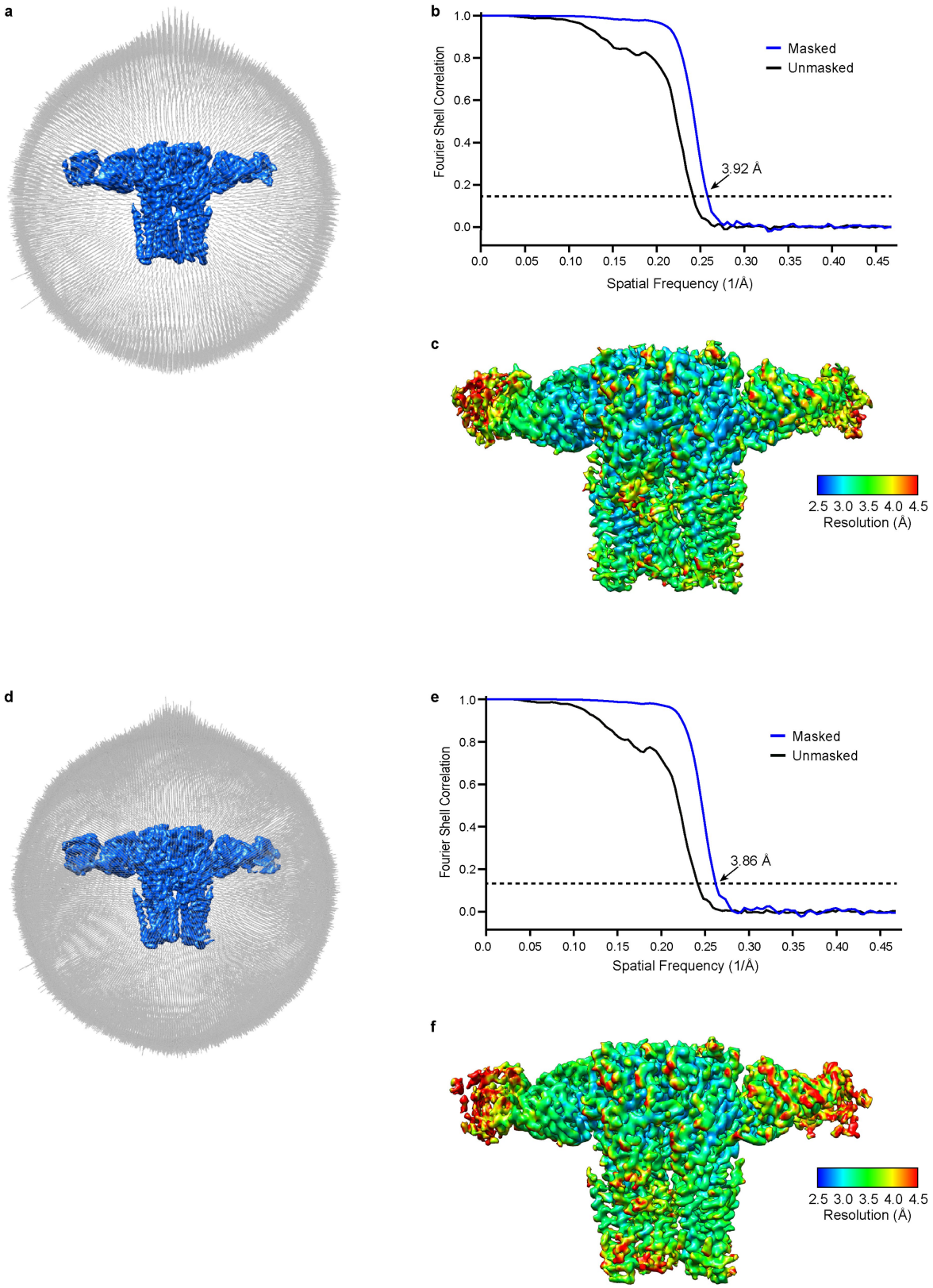
Extended Data Fig. 2 | Biochemistry and binding assay. a, FSEC of $GABA_A$ receptor with and without Fab bound, and SDS–PAGE analysis of a representative purification (from $n > 10$ purifications). **b**, Saturation binding assay with [^{3}H]-flumazenil. Single-site binding fits for receptor alone and receptor plus Fab both exhibited a Hill slope of ~1 (0.97 and 0.89, respectively). Plotted results are from a representative experiment performed in triplicate. $n = 3$ independent experiments. Data are shown as mean ± s.d. for a representative triplicate measurement. **c**, Competition of 10 nM [^{3}H]-flumazenil with diazepam. Calculated K_i for diazepam assumes a K_d for [^{3}H]-flumazenil of 7.7 nM. $n = 2$ independent experiments in triplicate. Data are shown as mean ± s.d. for a representative triplicate measurement. **d**, Dose–response experiments in the presence or absence of Fab. HEK cells were transfected with cryo-EM constructs and patch-clamped with or without pretreatment with 1 μM Fab for 1 min. Hill slopes are 1.7 and 1.4 with and without Fab, respectively. Published EC_{50} values for GABA range from 6.6 μM–107 μM[71–74]. $n = 3$ experiments from different cells. Data are plotted as mean ± s.d. **e**, Whole-cell patch-clamp recording of long application of cryo-EM ligands at concentrations used in cryo-EM samples to assess conformational state at equilibrium. The two traces shown are from one continuous recording; in between the two responses, Fab was added to 1 μM for one minute to saturate all receptor sites before second application of GABA and flumazenil (including Fab). $n = 3$ independent experiments. **f–g**, Docking of diazepam at the benzodiazepine-binding site based on superposition of benzodiazepine rings. The phenyl ring of diazepam would orient towards the membrane, possibly forming π–π-stacking interactions with Y58 on the complementary subunit. Similar to flumazenil, the halogen of diazepam could interact with H102, suggesting that this contact is conserved broadly among benzodiazepines and flumazenil. This orientation is largely consistent with predictions from a modelling and docking study[75], and distinct from that suggested by affinity labelling[76]. In this latter prediction, the diazepam phenyl group orients away from the membrane and would require local reorganization of side chains to avoid atomic clashes. **h–j**, Structural details of Fab–α1 interaction. Labelled residues are on the α-subunit. **i**, Top view. **j**, Side view.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 3 | Cryo-EM image processing procedure. **a**, Representative cryo-electron micrograph of the $GABA_A$ receptor–Fab complex. $n = 5{,}594$ images. **b**, Images of selected 2D classes from reference-free 2D classification by Relion. **c**, Overview of the image processing procedure (see Methods).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

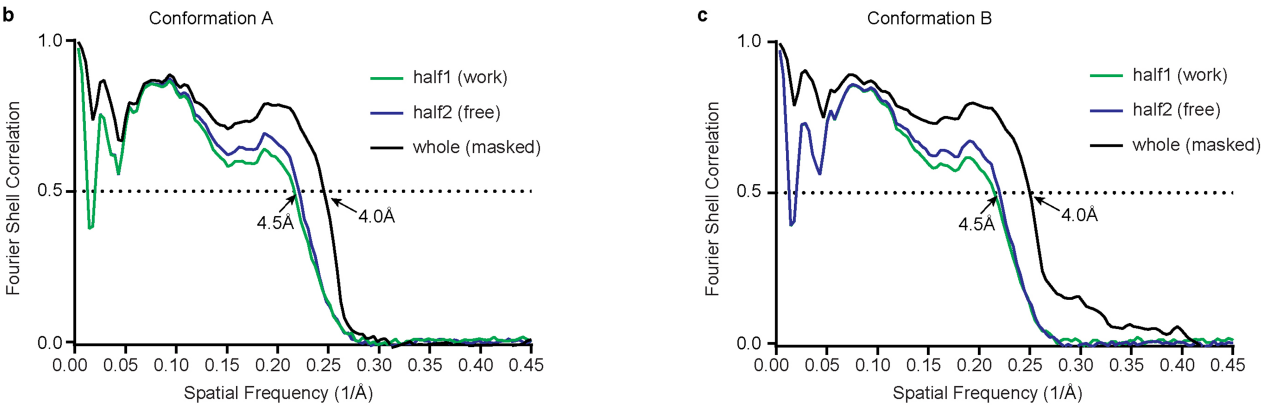


Extended Data Fig. 4 | Three-dimensional reconstructions of the two $GABA_A$ receptor conformations. a, Angular distribution histogram of $GABA_A$ receptor conformation A particle images. **b**, Fourier shell correlation (FSC) of conformation A maps before (black) and after (blue) masking. **c**, Local resolution of the $GABA_A$ receptor estimated by ResMap. **d**–**f**, as in **a**–**c** but for $GABA_A$ receptor conformation B.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

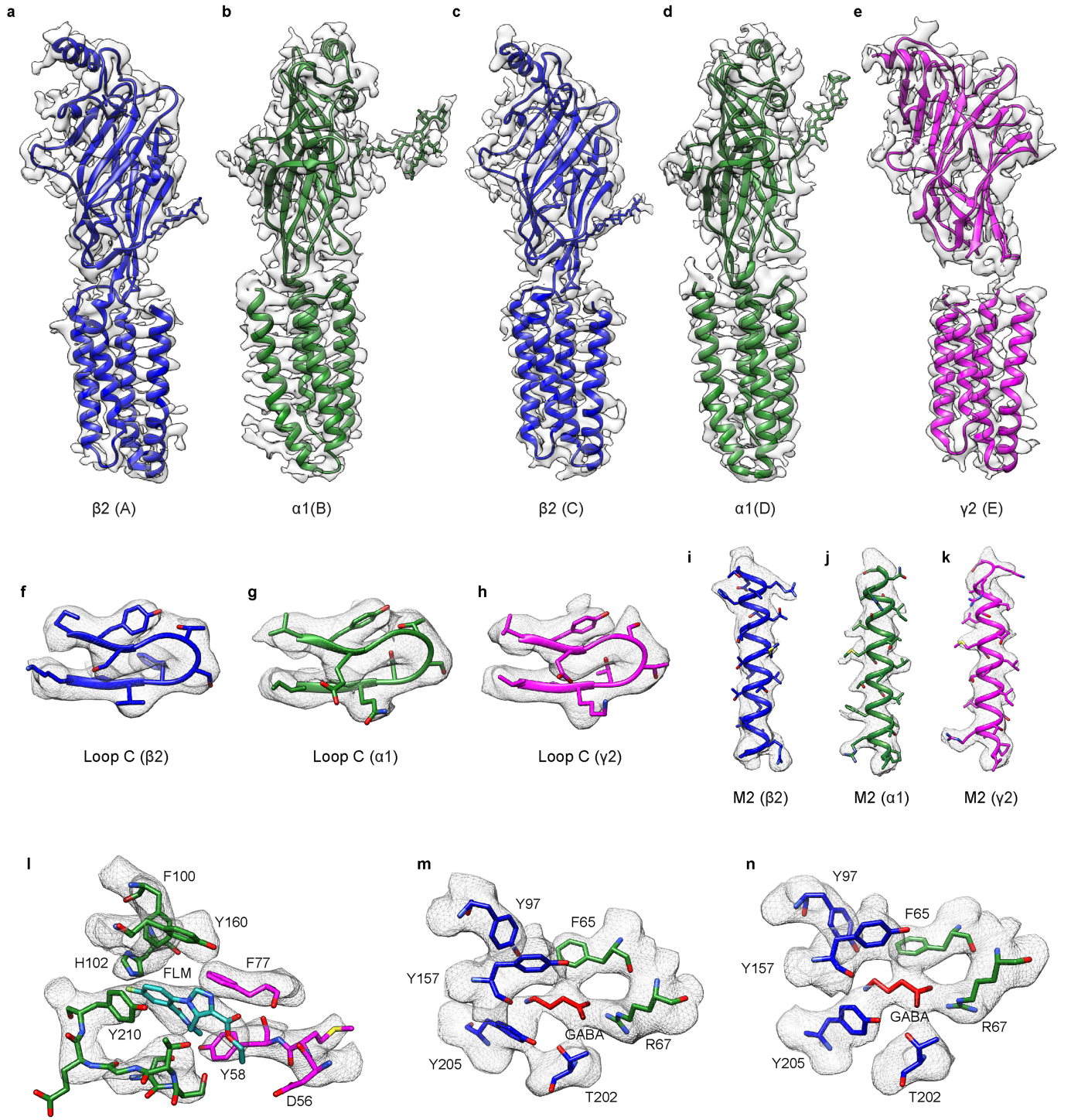
a

	Conformation A EMDB-7817 PDB 6D6U	Conformation B EMDB-7816 PDB 6D6T
Data collection		
Microscope	FEI Titan Krios	
Magnification	46,730	
Voltage (ke$^-$V)	300	
electron exposure (e$^-$/Å^{-2})	47	
Detector	Gatan K2 Summit with GIF energy filter	
Pixel size (Å/pixel)	1.07	
defocus range (μm)	-1.0 to -4.9	
Initial number of particles	1,050,737	
Final number of particles for reconstruction	292,662	200,442
Symmetry imposed	C1	C1
Box size (pixels)	260	260
Map Resolution (Å)	3.92	3.86
FSC threshold	0.143	0.143
Map-sharpening *B* factor (Å^2)	-186	-153
Refinement		
Number of non-H atoms	17,524	17,202
Protein residues	2,106	2,082
B factors (Å^2)		
Receptor	110	162
Fab	113	163
Glycosylation	105	163
GABA	69	118
Flumazenil	71	115
Cholesterol	155	217
Molprobity score	1.81 (100th percentile)	1.94 (100th percentile)
Clashscore	8.45 (97th percentile)	6.8 (100th percentile)
Poor rotamers (%)	0.48	2.65
r.m.s.d. values		
Bond lengths (Å)	0.005	0.004
Bond angles (Å)	0.72	0.67
Ramachandran analysis		
Favored (%)	95.0	96.3
Outliers (%)	0	0



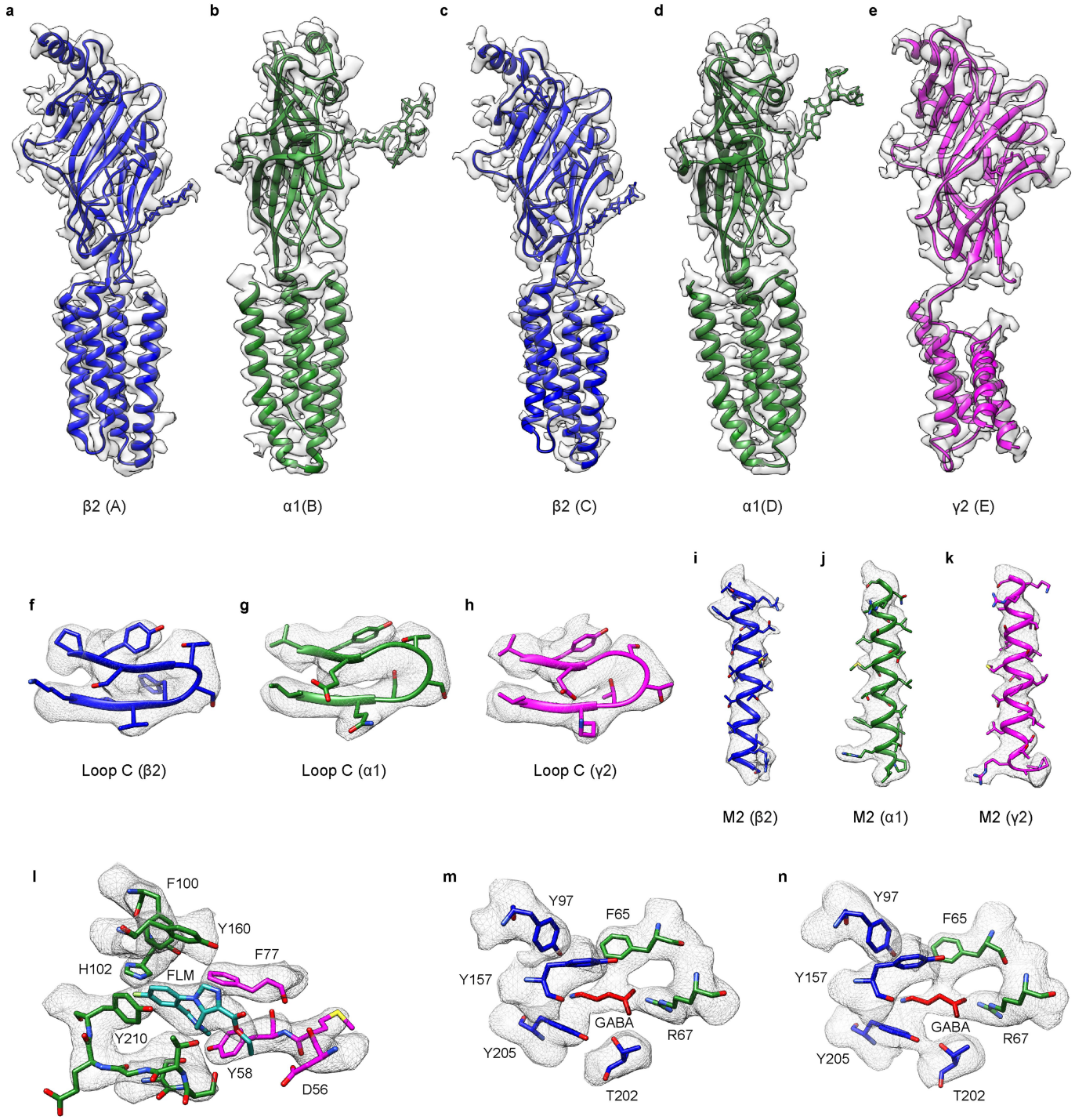
Extended Data Fig. 5 | GABA$_A$ receptor model map validation. **a**, Data collection and refinement statistics. **b**, **c**, FSC curves for cross-validation between the maps and models of both conformation A (**b**) and conformation B (**c**). FSC curves for final model versus summed map (whole) in black, for model versus half map in green (work), and for model versus half map not used for refinement in blue (free).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



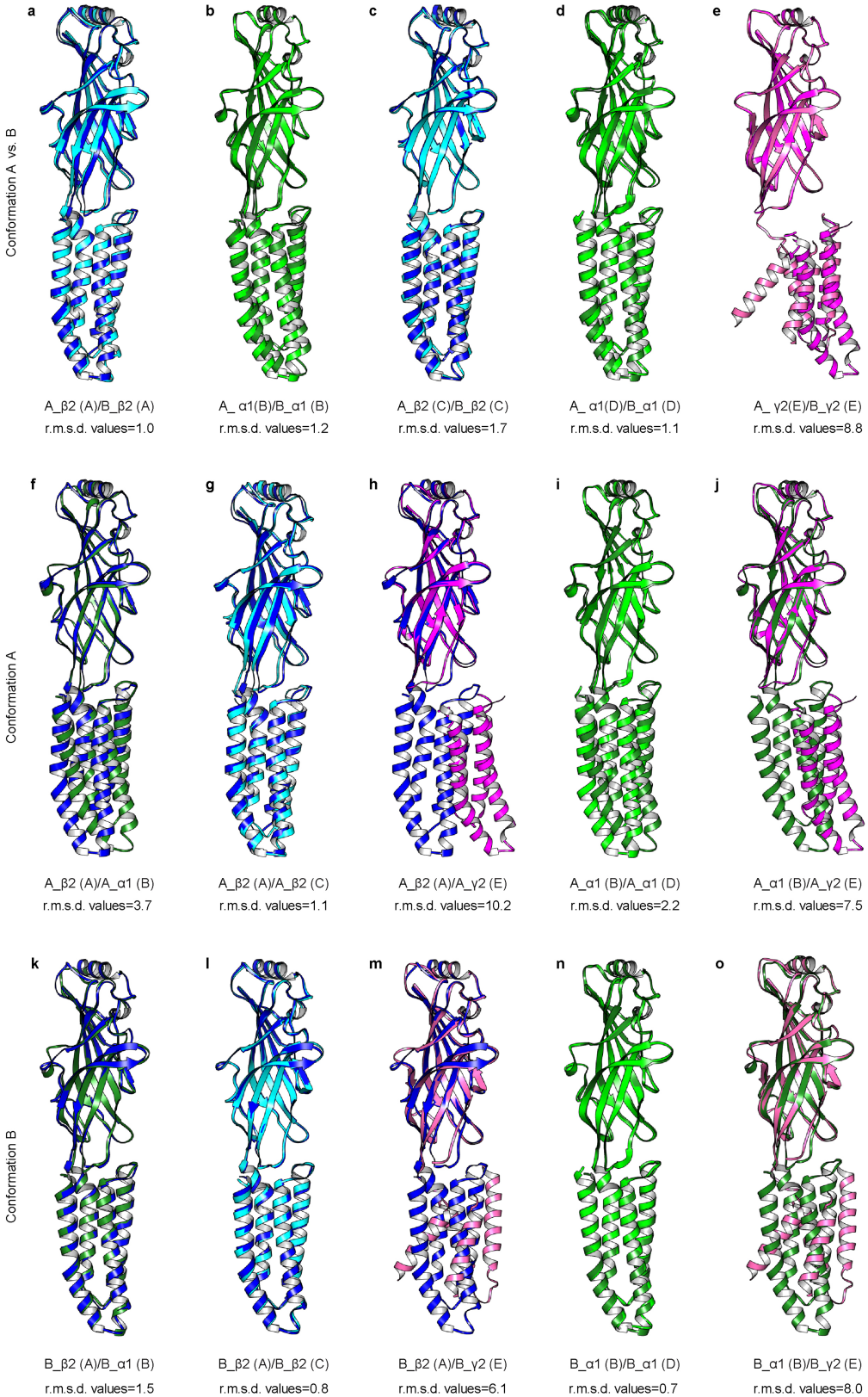
Extended Data Fig. 6 | Cryo-EM density of the $GABA_A$ receptor in conformation A. a–e, Cryo-EM density map of the $GABA_A$ receptor conformation A for a representative of each subunit. **f–h**, Cryo-EM density segments of loop C in α1, β2 and γ2-subunits. **i–k**, Cryo-EM density segments of M2 helix in α1, β2 and γ2-subunits. **l–n**, Cryo-EM density maps of ligand binding sites: flumazenil (**l**), two GABA binding sites (**m**, **n**).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



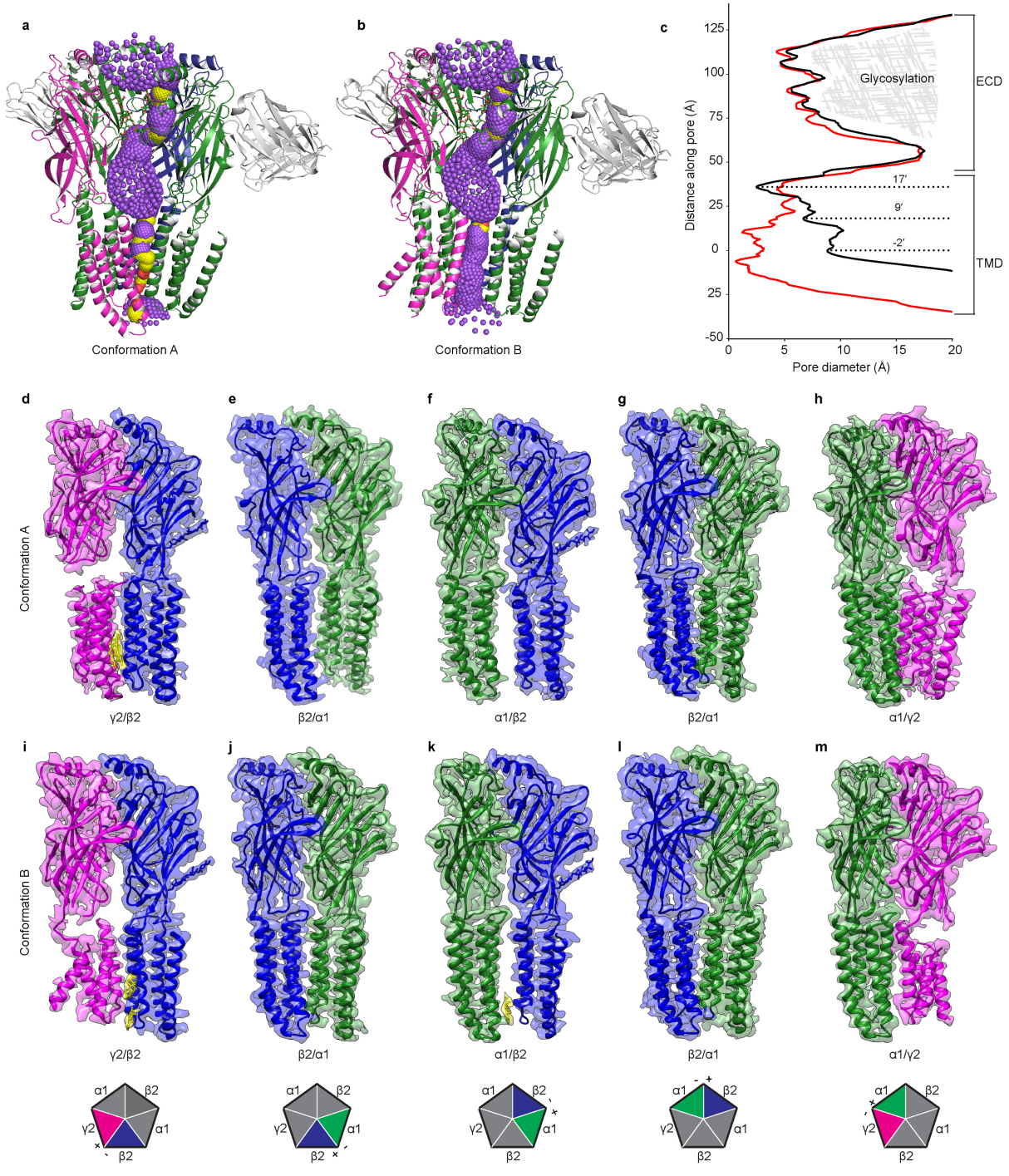
Extended Data Fig. 7 | Cryo-EM density of the $GABA_A$ receptor in conformation B. a–e, Cryo-EM density map of the $GABA_A$ receptor conformation B for a representative of each subunit; chain IDs are in parentheses. **f–h**, Cryo-EM density segments of loop C in α1, β2 and γ2-subunits. **i–k**, Cryo-EM density segments of M2 helix in α1, β2 and γ2-subunits. **l–n**, Cryo-EM density maps of ligand-binding sites: flumazenil (**l**), two GABA binding sites (**m**, **n**).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 8 | Superposition of subunits. a–e, Subunits of conformation A are compared to the corresponding subunit from conformation B. **f–j**, Superposition of subunits within conformation A. **k–o**, Superposition of subunits within conformation B. Chimera MatchMaker was used to generate alignments; r.m.s.d. values in Å are for Cα atoms over entire subunit. Chain IDs are in parentheses.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

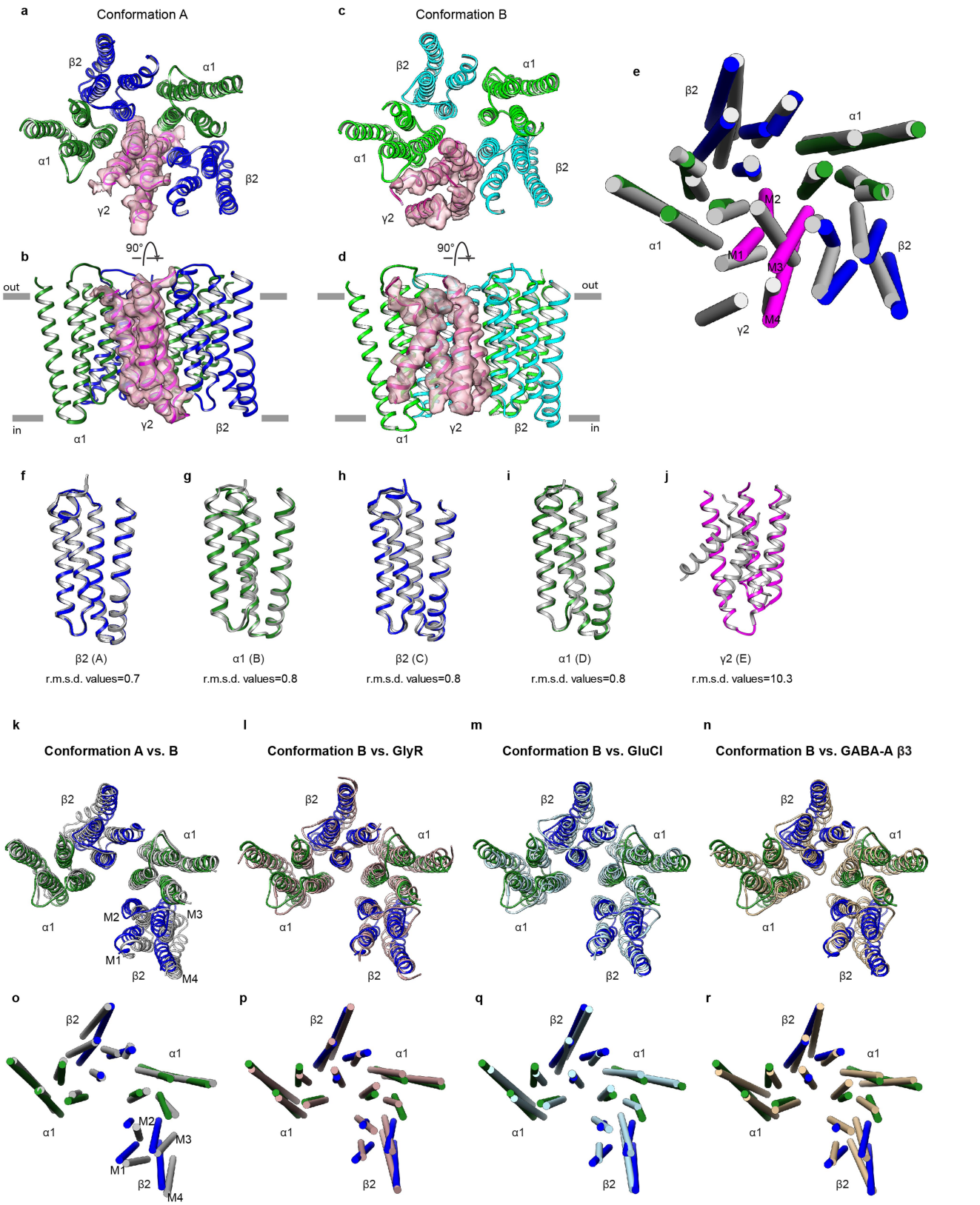


n

	Conformation A					Conformation B				
Interface	γ2/β2	β2/α1	α1/β2	β2/α1	α1/γ2	γ2/β2	β2/α1	α1/β2	β2/α1	α1/γ2
Interface area (Å^2)	2313	1968	1678	2643	1959	2074	2648	1699	2437	2272
Calculated Free Energy ΔG kcal/mol	-30	-22	-13	-44	-24	-28	-41	-14	-36	-30

Extended Data Fig. 9 | Permeation pathway and subunit interfaces. **a**, Cartoon of permeation pathway for conformation A. A single β2-subunit is removed for clarity. Purple spheres indicate pore diameters >5.6 Å; yellow is >2.8 Å and <5.6 Å; red is <2.8 Å. **b**, Same as **a** but for conformation B. **c**, Pore diameters for conformation A (red) and conformation B (black). The zero value along the *y* axis of the plot is aligned with the α-carbon of the −2′ position of conformation B. **d**–**m**, Side view of two adjacent subunits in conformations A (**d**–**h**) and B (**i**–**m**). The view is from the periphery of the receptor towards the pore axis. Cholesterol at the interface is shown in yellow in **d**, **i** and **k**. Cartoon pentagons (bottom) are coloured to illustrate all subunits composing the displayed interface; subunits not participating in the displayed interface are grey. Principal (+) and complementary (−) sides of the displayed interface are labelled on each pentagon. **n**, Analysis of the subunit interfaces of both conformations using PDBePISA server[69].

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 10 | Transmembrane domain flexibility and comparison with reference structures. a, **b**, Top and side view of the TMD of conformation A with density for the γ2-subunit shown. **c**, **d**, As in **a**, but for conformation B. **e**, Transmembrane domain superposition of conformation A (subunits in colour) over conformation B (grey). α-Helices are represented as cylinders. **f**–**j**, Superposition of single subunit TMD in conformation A (coloured) with its corresponding subunit in conformation B (grey). **k**–**r**, Superpositions of the four non-γ-subunits. Top and bottom rows contain the same superpositions in different representations. Conformation B is shown in all panels with α-subunits in green and β-subunits in blue. Reference structures include the glycine receptor with ivermectin bound (3JAF)[77], glutamate-gated chloride channel with ivermectin bound (3RHW)[78] and the $GABA_A$ β3 homopentamer (4COF)[16].

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

https://doi.org/10.1038/s41586-018-0265-1

Universality of free fall from the orbital motion of a pulsar in a stellar triple system

Anne M. Archibald[1,2]*, Nina V. Gusinskaia[1], Jason W. T. Hessels[1,2], Adam T. Deller[3,4], David L. Kaplan[5], Duncan R. Lorimer[6,7], Ryan S. Lynch[7,8], Scott M. Ransom[9] & Ingrid H. Stairs[10]

Einstein's theory of gravity—the general theory of relativity[1]—is based on the universality of free fall, which specifies that all objects accelerate identically in an external gravitational field. In contrast to almost all alternative theories of gravity[2], the strong equivalence principle of general relativity requires universality of free fall to apply even to bodies with strong self-gravity. Direct tests of this principle using Solar System bodies[3,4] are limited by the weak self-gravity of the bodies, and tests using pulsar–white-dwarf binaries[5,6] have been limited by the weak gravitational pull of the Milky Way. PSR J0337+1715 is a hierarchical system of three stars (a stellar triple system) in which a binary consisting of a millisecond radio pulsar and a white dwarf in a 1.6-day orbit is itself in a 327-day orbit with another white dwarf. This system permits a test that compares how the gravitational pull of the outer white dwarf affects the pulsar, which has strong self-gravity, and the inner white dwarf. Here we report that the accelerations of the pulsar and its nearby white-dwarf companion differ fractionally by no more than 2.6×10^{-6}. For a rough comparison, our limit on the strong-field Nordtvedt parameter, which measures violation of the universality of free fall, is a factor of ten smaller than that obtained from (weak-field) Solar System tests[3,4] and a factor of almost a thousand smaller than that obtained from other strong-field tests[5,6].

We observed PSR J0337+1715 with the Westerbork Synthesis Radio Telescope (WSRT), the Robert C. Byrd Green Bank Telescope (GBT) and the William E. Gordon telescope at the Arecibo Observatory (AO). We have more than 800 observations spanning approximately six years, which total about 1,200 h on-source. During each observation we folded (summed in time) the rotationally modulated radio signal from the pulsar according to a preliminary model for the its sky position, spin rate and orbital motion. We recorded flux density as a function of rotational phase, radio frequency and time. We processed these observations using standard techniques in precision pulsar timing[7] (Methods). We further averaged the data in time and frequency; most observations are averaged into roughly 20-min integrations with 20-MHz bandwidth. We compared each folded profile to a standard template (Extended Data Fig. 1) to determine how early or late the pulses arrived compared to our reference model. This resulted in approximately 27,000 multi-frequency pulse time-of-arrival measurements, with a formal weighted root-mean-square uncertainty of 1.0 μs (for individual telescope datasets: AO, 0.4 μs; GBT, 1.3 μs; WSRT, 1.6 μs).

To accommodate the complex three-body interactions in the system, we modelled the orbits by directly integrating the equations of motion[8]. To enable tests of the general theory of relativity, we include parameterized post-Newtonian[9] interactions between bodies. This framework allows essentially all gravitational theories to be approximated to first post-Newtonian order. If we forbid preferred-frame and preferred-location effects as well as non-conservation of momentum, then theories in this framework are parameterized by β, which measures the nonlinearity of gravity, and γ, which measures the degree to which gravity curves space-time. Both of these parameters take the value 1 in general relativity. We chose a point-particle Lagrangian that permits arbitrarily strong gravity internal to the bodies and parameterized post-Newtonian interactions between them[10]. We then used computer algebra[11] to construct equations of motion. Each orbit was specified by an initial system configuration at modified Julian date (MJD) 55,920.0 (2011 December 25 00:00:00 UTC). The evolution of this configuration was governed by β, γ and the strong equivalence principle (SEP)-violation parameter Δ. Because the self-gravity of the pulsar (which is a neutron star) exceeds that of the white dwarfs by a factor 10^4 and the SEP violation that we seek arises from self-gravity, we neglect possible SEP violations in the white dwarfs. Therefore, following ref. [12], we define $\Delta = m_G/m_I - 1$, the fractional difference between the inertial (m_I) and gravitational (m_G) masses of the pulsar. The SEP is satisfied only if $\Delta = 0$.

Our fitting procedure simulated orbits for trial sets of parameters. Once an orbit had been simulated, we used a linear least-squares fitting process to measure parameters such as the spin period of the pulsar and its offset from a reference sky position (Methods) and to obtain a goodness-of-fit parameter χ^2. We repeated this for many orbits to search the space of parameters for the best fit, residuals from which are shown in Fig. 1. We also computed numerical derivatives of the best-fitting orbit with respect to each parameter. This process gave us best-fit values and formal uncertainties on all parameters. We did not constrain the parameterized post-Newtonian parameters β and γ. By contrast, our observations and analysis procedure are able to constrain Δ.

The results of our fitting process are provided in Extended Data Tables 1–3. We measure $\Delta = -1.1 \times 10^{-6}$ with a formal 1σ uncertainty of 2×10^{-7}. We caution that the formal uncertainties that we provide do not include systematic effects such as excess delays caused by the variable solar wind (our line of sight to PSR J0337+1715 passes within 2.1° of the Sun every year) or refractive variations in the scattering time (roughly 30 ns). Such effects can be strongly correlated between measurements and can thus affect best-fit values substantially (and statistically significantly), despite being much smaller than the formal uncertainties on the pulse arrival times.

To obtain a realistic limit on Δ, we carried out a systematics analysis procedure on the residuals from our fit. The key idea was to look at the 'signature' of a non-zero Δ—that is, the effect on an orbit of introducing a non-zero Δ and then fitting for all other parameters (as in Fig. 1c). The differential acceleration introduced by a non-zero Δ shifts the orbit of the pulsar towards the outer companion. In the residuals from fitting all parameters except Δ, this effect produces a sinusoid with frequency $2f_{\mathrm{inner}} - f_{\mathrm{outer}}$ (where f_{inner} and f_{outer} are the inner and outer orbital frequencies, respectively; Methods). This signature differs from the $(f_{\mathrm{inner}} - f_{\mathrm{outer}})$-frequency behaviour that is sought in lunar laser-ranging experiments[4] because pulsar timing measures the line-of-sight distance,

[1]Anton Pannekoek Institute for Astronomy, University of Amsterdam, Amsterdam, The Netherlands. [2]Netherlands Institute for Radio Astronomy (ASTRON), Dwingeloo, The Netherlands. [3]Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, Victoria, Australia. [4]The Australian Research Council Centre of Excellence for Gravitational Wave Discovery (OzGrav), Hawthorn, Victoria, Australia. [5]Department of Physics, University of Wisconsin-Milwaukee, Milwaukee, WI, USA. [6]Department of Physics and Astronomy, West Virginia University, Morgantown, WV, USA. [7]Center for Gravitational Waves and Cosmology, Morgantown, WV, USA. [8]Green Bank Observatory, Green Bank, WV, USA. [9]National Radio Astronomy Observatory, Charlottesville, VA, USA. [10]Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, Canada. *e-mail: archibald@astron.nl

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

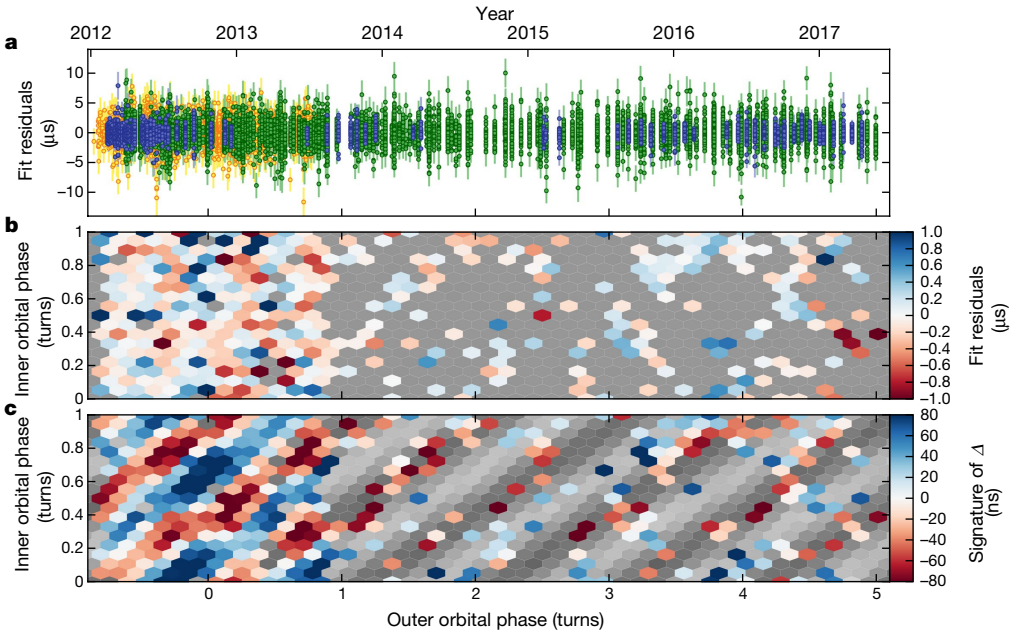


Fig. 1 | Difference between our pulsar time-of-arrival measurements and model (fit residuals). a, Residuals as a function of time. Orange points are WSRT data; green are GBT; blue are AO. Error bars are 1σ statistical uncertainties computed from the signal-to-noise ratio in our template-matching procedure. **b**, Values from **a** binned by inner orbital phase and outer orbital phase to reveal structure that varies with inner orbital phase. Axes are expressed in the (fractional) number of full turns that the corresponding orbit has taken. **c**, Signature of $\Delta = 2.6 \times 10^{-6}$: the pattern of residuals from an SEP-violating model after fitting a model with Δ fixed to zero. Blue and red show our actual observational sampling, whereas greyscale simulates uniform sampling. We claim that we would have detected the structure pictured in **c** if it had been present in **b**; note the different scales. Our systematics analysis procedure looks for structures in the residuals similar to that in **c** to quantify the presence of noise that might masquerade as an SEP violation.

so the signature of a non-zero Δ is additionally modulated by f_{inner} (Methods). We can also compute the signature of a non-zero Δ using our orbit simulator; see Fig. 1c. It is indeed a sinusoidal variation of pulse arrival times, with an amplitude of 30 ns when $\Delta = 10^{-6}$. The fact that the signature has a simple sinusoidal form suggests that, to understand how systematics might affect the measured value of Δ, we should look at sinusoids with similar frequencies $kf_{\mathrm{inner}} + lf_{\mathrm{outer}}$ (where k and l are integers). In Fig. 2 we show the results of fitting many such sinusoids to the residuals from our best fit. The distribution of coefficients of these sinusoids implies a 1σ scatter of 22 ns on the component that corresponds to the signature of a non-zero Δ (Methods). We thus obtain $\Delta = (-1.09 \pm 0.74) \times 10^{-6}$, which is consistent with

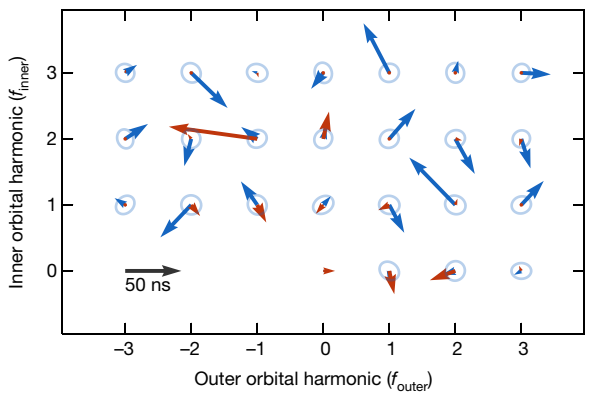


Fig. 2 | Quasi-Fourier representation of fit residuals. The length and direction of each arrow represents the amplitude and phase of a sinusoid, respectively. Each point represents a frequency that is the sum of a harmonic of the inner orbital frequency f_{inner} plus a harmonic of the outer orbital frequency f_{outer}. For reference, the length of the black arrow corresponds to an amplitude of 50 ns in the pulse arrival times for a given frequency. Blue arrows represent residuals from the fit. Ellipses represent 1σ arrow lengths determined from the formal uncertainties on our data; note that there are many signals present above this level. Red arrows represent the signature of an SEP violation of $\Delta = 2.6 \times 10^{-6}$. The longest red arrow—our limit on detectable signals—is 78 ns.

Table 1 | SEP-violation parameter Δ and uncertainties from various fitting approaches

	Δ ($\times 10^{-6}$)	Statistical uncertainty ($\times 10^{-6}$)	Amplitude of Δ signature (ns)	Systematic uncertainty (ns)
Primary fit				
Observatories: AO, GBT, WSRT Frequency band: $\nu_c \approx$ 1,400 MHz DM fitting interval: one year EOM: First-order PN, $\Delta \neq 0$	-1.1 ± 0.7	0.2	33	22
Alternative physical model				
EOM: Newtonian, $\Delta \neq 0$	0.7 ± 4.1	0.2	22	132
EOM: First-order PN of GR, $\Delta \equiv 0$	–	–	–	23
Subset of data				
AO (5,355 TOAs, WRMS = 0.6 μs)	-0.98 ± 2.5	0.3	25	64
GBT (18,487 TOAs, WRMS = 1.5 μs)	0.04 ± 1.2	0.4	1	35
WSRT (3,268 TOAs, WRMS = 1.8 μs)	-2.1 ± 2.5	1.3	47	55
Additional data				
$\nu_{c,AO} \approx$ 430 MHz, $\nu_{c,WSRT} \approx$ 350 MHz	-2.1 ± 1.5	0.2	67	47

The Δ constraint reported in the main text was obtained using the primary fit. Each other fit in the table differs in exactly one respect, by using an alternate physical model for the equation of motion (EOM), a subset of the data from the primary fit (from a single observatory) or additional data from different frequency bands for AO and WSRT. The second column gives the best-fit value for Δ with a 1σ uncertainty inferred from the systematics. The third column gives a 1σ formal statistical uncertainty. The fourth column gives the amplitude in nanoseconds of the signature of Δ that we found; the fifth column gives the 1σ amplitude of the systematics. DM, dispersion measure; PN, post-Newtonian approximation; GR, general relativity; TOAs, time-of-arrival measurements; WRMS, weighted root-mean-square of the residuals; ν_c, central observing frequency.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

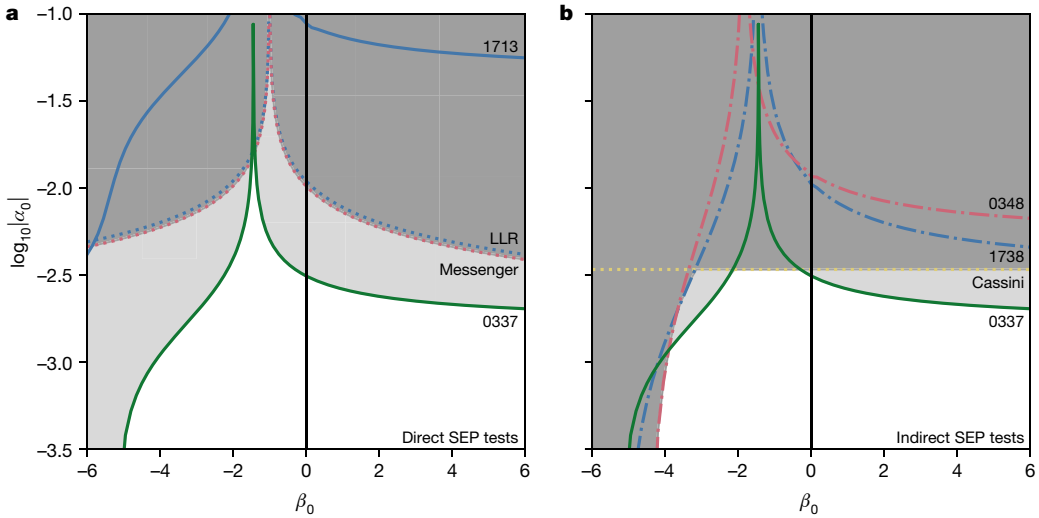


Fig. 3 | Constraints on quasi-Brans–Dicke theories of gravity. These theories are parameterized by α_0 and β_0. Existing constraints rule out the dark grey region. The constraint derived here, labelled '0337', additionally rules out the light grey region. We assume a very stiff equation of state for neutron stars ('.20' from ref. [25]) and are therefore conservative about neutron-star tests. **a**, Comparison with existing direct SEP tests: '1713', the wide binary PSR J1713+0747[6]; 'LLR', lunar laser ranging[4]; 'Messenger', Mercury-orbit[3] SEP tests. **b**, Comparison with indirect SEP tests: 'Cassini', Cassini Shapiro delay[20]; '1738' and '0348', dipole gravitational-wave limits from neutron-star–white-dwarf binaries[21,22].

zero at the 2σ level. This implies a 95%-confidence-level upper limit of $|\Delta| < 2.6 \times 10^{-6}$.

We explored the effect of fitting various subsets of our data and of handling physical effects in different ways. The values that we report in Extended Data Table 2 are calculated using the best method that we found; the results from this and other approaches are presented in Table 1. In each alternative approach, the limits that we obtain on Δ are compatible, but less constraining.

Our result is a direct test of the SEP: in the gravitational pull of the outer white dwarf, the pulsar and the inner white dwarf experience accelerations that differ fractionally by $|\Delta| < 2.6 \times 10^{-6}$. The most similar previous test is based on the (non-)evolution of the eccentricity of the pulsar–white dwarf binary PSR J1713+0747 falling in the gravitational pull of the Milky Way; this test yields $|\Delta| < 2 \times 10^{-3}$ (95% confidence) in a physically similar situation[6], but its sensitivity is limited by the low acceleration, which is 10^{-8} times that in our system. We are therefore able to improve on this previous limit by almost three orders of magnitude. Neutron-star experiments are particularly valuable because in many alternative theories of gravity the strong curvature of space-time inside the neutron star— among the largest accessible to observation[13]— would change its gravitational properties relative to those of less-compact objects. Among the many likely consequences of this change are SEP violation and the emission of dipolar gravitational waves.

To understand the theoretical implications of our result or to compare our result to other tests of the SEP, including indirect ones, we need to select a framework that parameterizes alternative theories of gravity. The parameterized post-Newtonian formalism[9] parameterizes most alternative theories to first post-Newtonian order. Although the interactions between bodies in the PSR J0337+1715 system are adequately described at this order, the interior of the pulsar is a strong-field region, and it is precisely this region that may cause SEP violations. The parameterized post-Newtonian framework is therefore insufficient to describe our result[14,15].

Although other strong-field frameworks are available[16] (Methods), we choose the family of quasi-Brans–Dicke theories[17]. These theories, inspired in part by Mach's principle[18], add a scalar field ϕ to general relativity. The gravitational constant G measured in a Cavendish experiment depends on the local value of ϕ. Within the family, α_0 and β_0 are parameters that select a particular theory (note that this β_0 is different from the parameterized post-Newtonian parameter β). Although Solar System experiments are able to constrain α_0, for large negative values of β_0 the phenomenon of 'spontaneous scalarization' occurs in neutron stars[19], allowing the value of ϕ inside them to be of order unity regardless of the weak-field behaviour. Quasi-Brans–Dicke theories are therefore well constrained by pulsar experiments; we summarize several key results, along with our constraint, in Fig. 3. In particular, this framework allows us to compare our own strong-field SEP test with the weak-field lunar laser-ranging test[4], a weak-field SEP test with MESSENGER[3], the weak-field light-bending test based on the Cassini mission[20] and pulsar tests that place upper limits on the emission of gravitational dipole radiation[21,22]. In the regime in which spontaneous scalarization does not occur (β_0 greater than about -4), our result provides the strongest upper limit on α_0 and hence the most stringent constraint on how gravity can deviate from the predictions of Einstein's general theory of relativity within the family of strong-field theories. We also improve markedly upon all other direct tests of the SEP.

Future results from the Gaia mission are expected to improve substantially on the light-bending limit obtained with Cassini[23], providing an improved indirect weak-field constraint. Although gravitational-wave observatories such as LIGO can carry out indirect strong-field SEP tests based on dipole gravitational radiation, even next-generation gravitational-wave detectors such as the Einstein Telescope and the Cosmic Explorer may not[24] improve on the direct SEP test obtained from PSR J0337+1715. On the other hand, detection templates for the Cosmic Explorer and Einstein Telescope need to involve dipole gravitational radiation only in the circumstances in which spontaneous scalarization might occur (when β_0 is less than about -4, the neutron star mass is in certain ranges, and certain equation-of-state assumptions are satisfied).

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0265-1.

Received: 12 April 2018; Accepted: 17 May 2018;
Published online 4 July 2018.

1. Einstein, A. Die Grundlage der allgemeinen Relativitätstheorie. *Ann. Phys.* **354**, 769–822 (1916).
2. Deruelle, N. Nordström's scalar theory of gravity and the equivalence principle. *Gen. Relativ. Gravit.* **43**, 3337–3354 (2011).
3. Genova, A. et al. Solar system expansion and strong equivalence principle as seen by the NASA MESSENGER mission. *Nat. Commun.* **9**, 289 (2018).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

4. Hofmann, F. & Müller, J. Relativistic tests with lunar laser ranging. *Class. Quantum Gravity* **35**, 035015 (2018).
5. Freire, P. C. C., Kramer, M. & Wex, N. Tests of the universality of free fall for strongly self-gravitating bodies with radio pulsars. *Class. Quantum Gravity* **29**, 184007 (2012).
6. Zhu, W. W. et al. Tests of gravitational symmetries with pulsar binary J1713+0747. Preprint at https://arxiv.org/abs/1802.09206 (2018).
7. The NANOGrav Collaboration. The NANOGrav nine-year data set: observations, arrival time measurements, and analysis of 37 millisecond pulsars. *Astrophys. J.* **813**, 65 (2015).
8. Ransom, S. M. et al. A millisecond pulsar in a stellar triple system. *Nature* **505**, 520–524 (2014).
9. Will, C. M. & Nordtvedt, K. Jr. Conservation laws and preferred frames in relativistic gravity. I. Preferred-frame theories and an extended PPN formalism. *Astrophys. J.* **177**, 757–774 (1972).
10. Nordtvedt, K. A post-Newtonian gravitational Lagrangian formalism for celestial body dynamics in metric gravity. *Astrophys. J.* **297**, 390–404 (1985).
11. Meurer, A. et al. Sympy: symbolic computing in python. *PeerJ Comput. Sci.* **3**, e103 (2017).
12. Damour, T. & Schaefer, G. New tests of the strong equivalence principle using binary-pulsar data. *Phys. Rev. Lett.* **66**, 2549–2552 (1991).
13. Baker, T., Psaltis, D. & Skordis, C. Linking tests of gravity on all scales: from the strong-field regime to cosmology. *Astrophys. J.* **802**, 63 (2015).
14. Will, C. M. in *Theory and Experiment in Gravitational Physics* Ch. 4 (Cambridge Univ. Press, Cambridge, 1981).
15. Damour, T. & Taylor, J. H. Strong-field tests of relativistic gravity and binary pulsars. *Phys. Rev. D* **45**, 1840–1868 (1992).
16. Berti, E. et al. Testing general relativity with present and future astrophysical observations. *Class. Quantum Gravity* **32**, 243001 (2015).
17. Damour, T. & Esposito-Farèse, G. Tensor-scalar gravity and binary-pulsar experiments. *Phys. Rev. D* **54**, 1474–1491 (1996).
18. Brans, C. & Dicke, R. H. Mach's principle and a relativistic theory of gravitation. *Phys. Rev.* **124**, 925–935 (1961).
19. Damour, T. & Esposito-Farese, G. Tensor-multi-scalar theories of gravitation. *Class. Quantum Gravity* **9**, 2093–2176 (1992).
20. Bertotti, B., less, L. & Tortora, P. A test of general relativity using radio links with the Cassini spacecraft. *Nature* **425**, 374–376 (2003).
21. Freire, P. C. C. et al. The relativistic pulsar–white dwarf binary PSR J1738+0333 – II. The most stringent test of scalar–tensor gravity. *Mon. Not. R. Astron. Soc.* **423**, 3328–3343 (2012).
22. Antoniadis, J. et al. A massive pulsar in a compact relativistic binary. *Science* **340**, 1233232 (2013).
23. Mignard, F. & Klioner, S. A. Gaia: relativistic modelling and testing. *Proc. Int. Astron. Union* **261**, 306–314 (2010).
24. Shao, L., Sennett, N., Buonanno, A., Kramer, M. & Wex, N. Constraining nonperturbative strong-field effects in scalar-tensor gravity by combining pulsar timing and laser-interferometer gravitational-wave detectors. *Phys. Rev. X* **7**, 041025 (2017).
25. Haensel, P., Proszynski, M. & Kutschera, M. Uncertainty in the saturation density of nuclear matter and neutron star models. *Astron. Astrophys.* **102**, 299–302 (1981).

Acknowledgements We thank P. Freire for pointing out how useful PSR J0337+1715 could be for testing the SEP, L. Shao for providing an independent cross-check on the signature of Δ and K. Nordvedt for explaining why the signature of Δ differs from that in lunar laser ranging. A.M.A. is supported by a Netherlands Foundation for Scientific Research (NWO) Veni grant. N.V.G. is supported by NOVA. J.W.T.H. acknowledges funding from an NWO Vidi fellowship and from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Starting Grant agreement number 337062 ('DRAGNET'). A.T.D. is the recipient of an Australian Research Council Future Fellowship (FT150100415). D.R.L. also received support from NSF award OIA-1458952. S.M.R. and I.H.S. are Senior Fellows of the Canadian Institute for Advanced Research. I.H.S. is also supported by an NSERC Discovery Grant. The NANOGrav project (involving D.L.K., D.R.L., R.S.L., S.M.R. and I.H.S.) receives support from National Science Foundation (NSF) Physics Frontiers Center award number 1430284. The National Radio Astronomy Observatory is a facility of the NSF operated under cooperative agreement by Associated Universities. The Arecibo Observatory is operated by SRI International under a cooperative agreement with the NSF (AST-1100968), and in alliance with Ana G. Mendez-Universidad Metropolitana and the Universities Space Research Association. The Green Bank Observatory is a facility of the NSF operated under cooperative agreement by Associated Universities. The WSRT is operated by ASTRON with contributions from NWO.

Reviewer information *Nature* thanks P. Freire and C. Will for their contribution to the peer review of this work.

Author contributions A.M.A. wrote the processing pipeline, orbital modelling, fitting and equation-of-state integration code, ran the data processing and fitting operations, wrote the manuscript, with substantial contributions from co-authors, and produced all figures and tables unless otherwise indicated. N.V.G. wrote the systematics analysis code, inspected observations for quality, carried out the systematics analysis, produced Table 1, Figs. 1, 2 and Extended Data Fig. 4, and wrote the section on orbital effects. J.W.T.H. carried out an intensive observing campaign with the WSRT. A.M.A., N.V.G., J.W.T.H., D.R.L., R.S.L., S.M.R. and I.H.S. carried out observations with Arecibo and the GBT. J.W.T.H., S.M.R. and I.H.S. carried out a preliminary version of the data processing. A.T.D. consulted on the astrometry. N.V.G. and D.L.K. carried out the tidal-effects analysis. All authors participated in discussions of the content of the paper.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0265-1.

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.M.A.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Precision timing. Arecibo observations were taken with the L-band Wide receiver, which has a dual linear polarization feed, and an 800-MHz band was recorded with the Puerto Rican Ultimate Pulsar Processing Instrument (PUPPI). GBT observations were taken with the L-band receiver, which has dual linear feeds, and an 800-MHz band was recorded with the Green Bank Ultimate Pulsar Processing Instrument[26] (GUPPI). WSRT observations were taken with the Multi-Frequency Front End receivers, which have dual linear feeds. A single tied-array beam on the sky was formed (using phase and polarization calibration determined by the observatory), and a 160-MHz band was recorded with the Pulsar Machine II[27] (PuMa II). All observations were coherently dedispersed[28].

Although we generally follow best practices developed by the pulsar timing array community[7], PSR J0337+1715 has a few unusual features that force us to adopt additional techniques. Unknown additional features, or known features that we were not able to compensate for completely, introduce systematic structure in our residuals. Our systematics analysis procedure serves to estimate their effect on the key parameter Δ, and our reported uncertainty includes the estimated effect of this systematic structure.

The full model describing the motion of the pulsar is too complicated to use in real-time observing. We therefore observed while folding 10-s integrations using the pulsar period predicted from a two-non-interacting-Keplerian model[8] (BTX) or a single-Keplerian-orbit model with varying parameters (BTX also) that is understood by the standard pulsar timing tool TEMPO. These simplified models predict pulse phases that can differ from the observed phases by a substantial fraction of a pulse period towards the end of our observing span. It is therefore necessary to correct the folded archives by phase-shifting the recorded profiles to match the predictions of the full model. This ensures that when we averaged archives into 20-min spans they were already aligned so that no further time smearing occurred. At the same time, by comparing the observing ephemeris to the full model, we were able to compute the phase drift of the observing model within each 10-s integration. It is impossible to correct for this smearing; its amplitude is typically around 200 ns and can be as large as 1,500 ns. More problematically, the model errors can easily be correlated with the inner or outer orbital phase, possibly in the same way as the signature of the SEP violation that we are looking for. This may explain some of the systematics that we detect, and we recommend that future observations be carried out with the more accurate short-term folding models that we currently use to realign archives. We plan to release a bundle of such short-term ephemerides covering at least the next few years.

The pulse profile from PSR J0337+1715 includes substantial linear polarization, varying as a function of pulse phase (Extended Data Fig. 1). Because all of our telescopes measure orthogonal pairs of polarizations directly, reconstructing the total intensity profile depends on accurate polarimetric calibration. The WSRT undergoes polarization calibration as part of the tied-array beam-forming process. Calibration data, including a feed and dish model, is available for the GBT, in addition to the noise diode scans that we took before each observation. Unfortunately, no feed and dish model is available for AO, and we found that despite our use of diode scans we were not able to calibrate AO polarimetry in some observations. Specifically, we have examples of AO observations where even with the best available calibration the reconstructed Stokes I (total intensity) profile differs substantially in shape from the standard template observation; in these cases a suitable polarization transformation is able to match the template to the observation. We therefore adopted a technique similar to 'matrix template matching'[29]: when we compare each observation to our accurately calibrated polarimetric template (shown in Extended Data Fig. 1), we fit for an arbitrary Mueller matrix, an offset in each of the Stokes parameters and a phase shift, transforming the template Stokes parameters to match the observed ones (the Stokes parameters summarize the polarization content of an averaged signal and consist of total intensity I, two components of linear polarization Q and U, and circular polarization V). This process makes our pulse arrival times largely insensitive to polarization calibration and also allows the polarization structure of the pulsar to constrain the timing, yielding a 15% improvement in fit uncertainties compared to a fit using only the total intensity.

Standard practice in precision pulsar timing is to take low-frequency (for example, 430 MHz) observations quasi-simultaneously with each high-frequency (typically 1,400 MHz) observation to better constrain dispersion-measure variations. Our AO observations were taken in this mode, and on some days WSRT data were acquired at 350 MHz. Unfortunately, we found that if we included this low-frequency data (using a template based on a bright AO 430-MHz observation) our estimate of the systematics in the post-fit timing residuals became measurably worse. This may be the result of interstellar scintillation and scattering: at 1,400 MHz, we observe scintillation with a typical frequency structure of 5 MHz. This predicts[30,31] a scattering tail of 30 ns, varying by a factor of roughly two on timescales of months owing to refractive scintillation. This is comparable to the size of the signals that we are looking for. Although such scintillation is a minor systematic effect on the 1,400-MHz data, the scattering timescale is predicted to increase as the negative fourth power of observing frequency[31], giving a scattering timescale of roughly 8 μs at 430 MHz, large enough to complicate our use of low-frequency data. We therefore omitted these low-frequency observations in our primary fit (but see Table 1 for an evaluation of their effect on our result when included).

Our observations primarily record frequencies of 1,100–1,900 MHz. We expect the intrinsic profile of the pulsar to vary as a function of frequency across this range. Nevertheless, we use the single pulse profile template shown in Extended Data Fig. 1 for all observations in this band. We therefore expect there to be a modest frequency-dependent but time-independent time shift in our data. To compensate for this, we fix the dispersion measure and fit for a delay that is a polynomial function of the logarithm of frequency[7]. Using the F test we found that four terms were sufficient to model this variability. We therefore include four parameters in our timing model to describe this frequency variation.

Because the ecliptic latitude of PSR J0337+1715 is only 2.1°, every March our line of sight to the pulsar passes very close to the Sun. The solar wind then contributes potentially substantial extra dispersion measure to these observations. Although we fit for an idealized solar-wind model each year (see below), we know that the solar wind is time-variable and not spherically symmetric. We therefore excise all data for which the line of sight passed within 5° of the Sun; this keeps the predicted excess delays due to the solar wind below a few microseconds. Our solar-wind fitting should remove the majority of this, and our systematics estimate should account for the residual effects on our estimate of Δ.

Finally, in this very large collection of 818 observations, a few will inevitably have been corrupted by observer error, telescope malfunctions or radio-frequency interference. We therefore constructed a summary plot for each observation showing pulse profile versus time and frequency, smearing within an observing subintegration, and timing residuals relative to the short-term ephemeris used to align the observation. We examined these by eye so that we could excise part or all of any problematic observation. In addition to the standard automatic interference excision provided by the program *paz* from PSRCHIVE, we found it necessary to manually excise interference from 65 observations and to completely discard 17.

Timing model. Traditional pulsar timing models rely on formulae expressing Keplerian, or parameterized post-Keplerian, orbits. No such formulae are known that can handle the classical three-body interactions that occur in this system. We therefore implemented our timing model by directly integrating the equations of motion with a Bulirsch–Stoer integrator[32], using root-finding methods on the integrator's dense output to compute the pulsar proper time at which each received pulse was emitted. This orbital modelling has only finite accuracy, limited both by the step size of the differential-equation integrator and by the numerical precision with which the millions of steps are accumulated. We addressed truncation error by using an adaptive-step-size integrator with a tolerance parameter; we adjusted this tolerance parameter to obtain a negligible truncation error of approximately 0.1 ns (Extended Data Fig. 2). To obtain sufficiently small round-off error, we used 80-bit floating point to carry out the integrations. We verified that 80-bit precision was adequate by cross-checking against an implementation using 128-bit floating-point routines, which ran 50 times slower but agreed with the 80-bit results to sufficient accuracy.

The parameters describing a hypothetical timing solution fall into two categories. Some parameters, such as the outer binary period, affect the orbit of the pulsar, requiring a new orbit to be simulated when they are changed. Other parameters, such as the spin frequency of the pulsar, can be determined by a linear least-squares fit once the nonlinear parameters have been set. Because simulating an orbit takes roughly 60 s, and because there are strong covariances between some parameters in the two categories, we carry out the optimization in two nested stages, one a nonlinear downhill optimizer and the other a simple linear least-squares solver. For Bayesian computations, we are able to operate on the nonlinear parameters alone by analytically marginalizing over the linear parameters. This analytical marginalization amounts to using the linear least-squares best-fit values for the linear parameters and adding a correction to the log-probability computed from the linear least-squares fit matrices. The linear (or approximately linear to high accuracy) parameters are as follows. To parameterize the spin of the pulsar, we fit for pulsar spin frequency and frequency derivative. For astrometric parameters, we begin with published values of position and distance[8], set proper motion to zero and then fit for offsets from these values as linear parameters: position offsets, parallax error and proper motion. We also fit for instrumental delays between telescopes and for a time-independent but frequency-dependent delay due to profile variations with frequency (see above). Finally, to accommodate variations in the dispersion measure to the pulsar, we fit for one dispersion measure value per year (interpolating linearly between these values) and for an interplanetary medium delay (changing as our line of sight passes through different parts of the Solar System; SOLARN0 in TEMPO2[33]) each year.

Our orbital simulations operate on pulse arrival times that have been 'barycentred' by TEMPO2: arrival times measured at an observatory are converted to

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

a suitable relativistic timescale (coordinate barycentric time, TCB), corrected for interstellar medium delays and converted to times the pulse would have arrived at the Solar System barycentre. TEMPO2 provides derivatives of these arrival times with respect to astrometric parameters; we use these derivatives in the linear part of the fitting (a reasonable approximation because the delays associated with the fitted values are well below one millisecond). This ensures that the orbital simulator can operate without knowledge of the pulsar's motion in the plane of the sky or the Earth's motion in the Solar System. However, such an approach necessarily ignores various Kopeikin effects[34], where our view of the system interacts with orbital fitting. For example, the transverse motion of PSR J0337+1715 in the plane of the sky and the orbital motion of the Earth both change the angle from which we view the pulsar's orbit at any given time, albeit very slightly. We used a sample orbit and a timing formula that correctly combines Earth and pulsar position to confirm that these effects are too small to be important, and also that they are confined to relatively long timescales (one year, one outer orbit, the span of all our observations) that far exceed the approximately 0.8-day timescale of the signature of Δ.

The reduced-dimensionality fitting problem for the nonlinear parameters has some strong covariances (Extended Data Fig. 3) despite our attempt to choose a natural parameterization of the orbit. Nevertheless, the posterior distribution seems to be multivariate normal, so Bayesian methods should agree with simpler frequentist calculations.

Fit quality and systematics. To evaluate the quality of the fit, we investigate the distribution of timing residuals. Specifically, we consider the distribution of the residuals divided by the formal uncertainties on the corresponding data (Extended Data Fig. 4). In an ideal situation this distribution should be Gaussian with zero mean and unit standard deviation; in our data the actual scatter exceeds the claimed uncertainties by a factor of 1.164. Although it is common practice in pulsar timing to rescale the formal uncertainties for each telescope so that $\sigma = 1$, in our data there are strong correlations between residuals, which motivates a more sophisticated approach to systematics.

Once we have selected a best-fit solution, we compute derivatives of the pulse arrival times with respect to all parameters of the solution. For the linear parameters, no additional computation is necessary, but for the parameters that affect the orbit we compute numerical derivatives using the Python package *numdifftools*. With this set of partial derivatives we compute the signature of Δ: we take the partial derivative with respect to Δ and then least-squares fit and subtract the partial derivatives with respect to all other parameters. This produces the structure in the pulse arrival times that is uniquely explainable by a change in Δ. Examining this signature, we find that it is nearly sinusoidal, with a frequency of $2f_{\rm inner} - f_{\rm outer}$ and a specific phase. The amplitude of this sinusoid also provides a conversion factor between amplitude in nanoseconds and size of Δ (which is dimensionless): if $\Delta = 10^{-6}$, then the signature will be a sinusoidal variation in pulse arrival times with an amplitude of 30 ns.

Why should the signature of an SEP violation be a sinusoid with frequency $2f_{\rm inner} - f_{\rm outer}$? In lunar laser-ranging experiments[4,35], the signature of an SEP violation is an offset of the inner orbit in the direction of the outer companion; in Earth–Moon ranging data this should be a signal with frequency $f_{\rm inner} - f_{\rm outer}$, and indeed the same results are obtained when searching for this sinusoid in the residuals from a general-relativity-obeying fit as when introducing an SEP violation into the physics being integrated[4]. Our observations differ in a key way from the lunar laser-ranging observations: whereas lunar laser ranging measures the Earth–Moon distance directly, in the PSR J0337+1715 system we measure the line-of-sight distance to the pulsar but not (directly) the distance between the pulsar and its inner companion. If, because of an SEP violation, the inner binary separation varies with a frequency $f_{\rm inner} - f_{\rm outer}$ and the pulsar orbits the inner centre of mass with frequency $f_{\rm inner}$, then our measurements of line-of-sight distance vary as the product of sinusoids at these two frequencies. Therefore, an SEP violation should produce sinusoids at frequencies of both $2f_{\rm inner} - f_{\rm outer}$ and $f_{\rm outer}$. The latter frequency can readily be absorbed into fitting the orbital parameters, which are not known a priori, but the former frequency cannot; it is the unique signature of an SEP violation.

We were concerned about systematics that affected our measurement of Δ. The signature computed above shows what structure a systematic should have to influence Δ: a sinusoid of frequency $2f_{\rm inner} - f_{\rm outer}$. At this particular frequency we cannot distinguish directly between systematics and genuine physical deviations from general relativity. We therefore look at a collection of harmonically related frequencies, most of which are not associated with known physical effects related to the motion of PSR J0337+1715. Specifically, we look at frequencies $kf_{\rm inner} + lf_{\rm outer}$ for modest integer values of k and l. Given any collection of residuals, we compute a quasi-Fourier representation by least-squares fitting a family of sinusoids to the residuals. Each sinusoid is of the form:

$$h_{k,l}(t) = \mathrm{Re}(C_{k,l} e^{2\pi i (k f_{\rm inner} + l f_{\rm outer}) t})$$

This is not exactly a two-dimensional Fourier series because our data are unevenly sampled, but it resembles such a representation and the sinusoids are approximately orthogonal. In Fig. 2 we show such a quasi-Fourier representation of the residuals from our primary fit, as well as a quasi-Fourier representation of the signature of Δ. Most quasi-Fourier coefficients exceed the formal uncertainties (ellipses), indicating that there is structure in our data at these frequencies. Other quasi-Fourier coefficients, related directly to modelled physical effects, are nearly zero because the fitting procedure has removed most or all of the power at those frequencies. The signature of Δ (red arrow) appears almost entirely in a single quasi-Fourier coefficient, so if we can estimate the typical sizes of quasi-Fourier coefficients due to systematics then we can infer the systematic contribution to Δ.

To estimate the typical systematic contribution, we assume that all of the Fourier coefficients are drawn from the same normal distribution, with zero mean. If we could compute the standard deviation of this distribution, then we would know the probability distribution of the quasi-Fourier coefficient that looks like the signature of Δ; we could then infer the systematic uncertainty on our estimate of Δ.

The challenge in working with the Fourier coefficients is that the fitting process unavoidably removes power from some of them. For example, the coefficients at $f_{\rm inner}$ are removed by fitting for $a\sin i_{\rm I}$ (the projected semi-major axis of the inner orbit) and $T_{\rm asc,I}$ (the time of the ascending node of the inner orbit). The full nonlinear fitting process is computationally expensive and requires manual intervention, so we rely on the numerical derivatives that we computed earlier to carry out a linear approximation to our fitting procedure. We therefore repeatedly generated synthetic sets of residuals with systematics drawn from a normal distribution with unit standard deviation. For each synthetic set of residuals, we (linear least-squares) fit and remove the derivatives with respect to all parameters. We then compute the power in the remaining Fourier coefficients and scale the synthetic dataset so that the remaining power matches that in the real dataset. During the fitting we also obtain the systematic contribution to Δ (appropriately rescaled). We collect these Δ values from 10^5 synthetic datasets and then use this distribution to obtain 1σ uncertainties and 95% upper limits on Δ.

Timing solution summary. We divide the system parameters into three categories: those we fix in the fitting (Extended Data Table 1), those we fit for directly (Extended Data Table 2) and those we infer from the fit parameters (taking covariances into account; Extended Data Table 3).

Although our fit allows astrometric parameters to vary, these parameters have strong covariances because the ecliptic latitude of the system is only 2.1° and are readily affected by year-long systematics such as uncorrected interplanetary medium effects. We therefore recommend against using the values quoted here for astrometric purposes. In an upcoming work we plan to compare astrometry derived from pulsar timing with that obtained from a very-long-baseline interferometry campaign.

Orbital effects. In our timing model we use first-order post-Newtonian equations of motion of three point particles; that is, we neglect orbital effects caused by tidal deformation of the stars and higher-order post-Newtonian effects such as frame dragging[36] and gravitational wave emission[37]. These effects fall off strongly with the distance between objects and, given that even the inner binary of the triple system has a relatively wide orbit (16 light-seconds), are too small to affect our data. Here we estimate the impact of these effects on the measured orbital parameters.

Periastron advance. The first-post-Newtonian-order relativistic periastron advance and classical periastron advance caused by three-body interactions are included in the fit, because they are taken into account in the equations of motion. Some additional periastron advance can be caused by tidal deformation of the inner white-dwarf companion of the system[38] or by any higher-order post-Newtonian effects[1,36]. We calculate the change in the longitude of periastron of the pulsar orbit caused by tidal asymmetry of the companion[39,40] assuming a tidal Love number of $k_2 \approx 0.01$, appropriate for a helium-core white dwarf with an extended envelope[41]: $\dot{\omega}_{\rm tidal} = 3 \times 10^{-5}$°yr^{-1}. This value is about five orders of magnitude smaller than relativistic periastron advance[1] $\dot{\omega}_{\rm rel} = 0.12$°yr^{-1} and about three orders of magnitude smaller than what can be detected with the current precision of the existing data (see Extended Data Table 2). At higher post-Newtonian order, the periastron advance of the inner orbit is dominated by the Lense–Thirring effect, in which interaction of the orbital angular momentum and the spin of the inner white dwarf causes the orbital plane to precess[42]. We estimate this precession to be no greater than $\Omega_{\rm SO} = 3 \times 10^{-4}$°yr^{-1}, even if the white dwarf is rotating as rapidly as once per minute and its angular momentum is aligned with the angular momentum of the inner orbit.

Dissipative effects. The tidal deformation of the inner white dwarf can cause a loss of energy from the system (tidal lag), thereby shrinking the orbit. Assuming an effective tidal parameter of $Q \approx 10^7$ for the inner white dwarf[41,43], we calculate the characteristic timescale of the decrease in the orbital period[39] due to this effect to be $\tau_{\rm p} \equiv P_{\rm orb} / \dot{P}_{\rm orb} \approx 2 \times 10^{17}$ yr. The characteristic timescale of the orbital decay due to gravitational-wave emission is $\tau_{\rm gr} \approx 2 \times 10^{12}$ yr. The current precision with which we can measure $P_{\rm orb}$ for the inner binary is approximately 10^{-8} days.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

This means that we need about 100 years of observations to be able to detect dissipation of the orbit of the inner binary due to gravitational-wave emission and about 10^4 years to detect the dissipation due to tidal deceleration.

Theoretical implications. Our central result is that, in the same gravitational field, the fractional difference in accelerations $|\Delta|$ between a 1.4359-solar-mass pulsar and a white dwarf is no more than 2.6×10^{-6} (at 95% confidence). We can compare this directly to a previous test in which the pulsar PSR J1713+0747 and its white-dwarf companion were observed falling in the Galactic potential[6]; this test constrains $|\Delta|$ to less than 2×10^{-3} (95% confidence), so our result is an improvement by three orders of magnitude. That said, we would like to compare our result to the weak-field SEP test carried out by lunar laser ranging or to indirect limits on SEP violations that come from upper bounds on gravitational dipole radiation. Such comparisons require a theoretical framework.

The logical structure of equivalence principles that lead to general relativity is complicated[44], but the key idea here is the relationship between the universality of free fall and the SEP. The universality of free fall is something that we can demand of any theory of gravity, whether Newtonian or post-Newtonian. It is frequently restricted to apply only to objects without strong self-gravity, because the nonlinear superposition of gravity in post-Newtonian theories could easily lead to violations of the universality of free fall in cases where the strong self-gravity of an object interacts with some external gravitational field. The SEP extends this requirement to cover systems with strong self-gravity and places additional requirements on local position invariance and local Lorentz invariance—all of which are also required to apply to systems with strong self-gravity. The complete SEP poses a very stringent constraint on the form that the nonlinearity of gravity can take, sufficiently so that very few theories satisfy the full SEP[2]. Our result thus tests the intersection of these two principles: we test the universality of free fall in the SEP regime in which bodies have potentially significant self-gravity. Further, in our system only one of the objects, the pulsar, has gravity firmly in the post-Newtonian regime. This permits violations of the SEP and the universality of free fall to be parameterized simply by allowing the gravitational and inertial masses of the pulsar to differ[10].

Lunar laser ranging places a limit of $|\Delta| < 1.3 \times 10^{-14}$ (95% confidence) on the Earth–Moon–Sun system[4]. Because this system is well approximated by the first post-Newtonian order, we can describe SEP violations by the (weak-field) Nordtvedt parameter η_N: $\Delta = \eta_N E_B$, where E_B is the fractional gravitational binding energy of the test body. What is actually measured is the difference in acceleration between the Earth and the Moon; both bodies would experience an SEP violation, but E_B for the Moon is about 1/25 times that for the Earth, so we incorporate it alongside that of the Earth as a correction to E_B. Because the difference in E_B between the Earth and the Moon is -4.45×10^{-10}, lunar laser ranging constrains[4] $|\eta_N|$ to less than 2.4×10^{-4}. If we simply apply this to the PSR J0337+1715 system, then $E_B \approx 0.1$ and $|\Delta| < 2.6 \times 10^{-6}$; consequently, the (strong-field) Nordtvedt parameter is $|\hat{\eta}_N| < 2.6 \times 10^{-5}$ and we appear to improve on the lunar laser-ranging result substantially. However, the phenomenon of (potential) SEP violation arises from the interior of the pulsar, where the first post-Newtonian order is an insufficient approximation. Directly comparing weak-field η_N and strong-field $\hat{\eta}_N$ depends on having a strong-field theory of gravity. Unfortunately, no completely general framework exists for describing strong-field effects, so it is necessary to specialize to specific families of strong-field theories of gravity. An overview of the space of possibilities is provided in ref. [16].

Within the context of tensor–(multi-)scalar theories, there is a parameterization of the second post-Newtonian order that involves only four parameters[45]: the β and γ parameters of standard parameterized post-Newtonian models, along with ε and ζ, which describe the second-post-Newtonian-order effects. Combining the lunar laser-ranging constraint on $\eta_N = 4\beta - \gamma - 1$ with our constraint on Δ, we infer a limit of $|\varepsilon/2 + \zeta| < 10^{-3}$. Nevertheless, in the interior of a neutron star, the second-post-Newtonian-order approximation may not be sufficient either[14,15].

It has been suggested[46] that pulsar timing results could be described in terms of constraints on scalar coupling constants α_j for the bodies involved. The values for these coupling constants depend on the scalar–tensor theory being considered and the equation of state assumed for the neutron star. Expressing the constraints in this way gathers almost all the theory and equation-of-state dependence in these α_j parameters, providing a nearly theory-independent way of comparing pulsar timing results. In our case, this is straightforward, because $|\Delta| = |\alpha_o(\alpha_p - \alpha_i)| < 2.6 \times 10^{-6}$. Here α_i and α_o are coupling constants for the inner and outer white dwarfs, respectively, equal to a weak-field coupling constant in many theories, and α_p is the scalar coupling constant for this pulsar, which has a mass of 1.4359 solar masses. Limits on dipole gravitational radiation constrain $(\alpha_p - \alpha_i)^2$, implying a different dependence on theory parameters. Some theories, including TeVeS[47], predict that $\alpha_p = \alpha_i = \alpha_o$; therefore, these theories cannot be constrained by our result[21,48].

A family of tensor–multi-scalar theories with first-post-Newtonian-order terms that agree exactly with general relativity but that have strong-field behaviour parameterized by β' and β'' has been introduced previously[19]. Our Δ measurement implies $|\beta'| < 3.5 \times 10^{-3}$; combining this with existing results[49] also implies $|\beta''| < 1$. However, this family of theories has serious theoretical problems, including the presence of negative-energy excitations.

The standard framework for comparing strong-field tests of general relativity involves the quasi-Brans–Dicke theories[17], which are parameterized by α_0 and β_0. Standard Brans–Dicke gravity arises in the special case $\beta_0 = 0$, in which the Brans–Dicke parameter ω_{BD} is related to α_0 by $\alpha_0^2 = (2\omega_{BD} + 3)^{-1}$. Quasi-Brans–Dicke theories can, under certain circumstances, arise as local approximations to theories in which the scalar has a potential that causes it to vary on cosmological scales[17]. In Solar System tests, the relevant quantity is $|\alpha_0|$, and by making this small enough weak-field deviations from general relativity can be suppressed enough to pass any given test. Assuming standard Brans–Dicke gravity, the Cassini Shapiro delay measurement[20] requires ω_{BD} to be greater than about 15,000; the constraint that we derive implies that ω_{BD} is greater than about 73,000 (95% confidence). However, assuming quasi-Brans–Dicke gravity, for sufficiently negative values of β_0 massive pulsars can undergo 'spontaneous scalarization' and acquire an order-unity deviation from general relativity regardless of the weak-field behaviour of the theory. In other words, regardless of how well Solar System tests constrain the weak-field behaviour of gravity, this family of theories can still exhibit strong-field behaviour that differs substantially from general relativity.

Using this quasi-Brans–Dicke family of theories, combined with a neutron-star equation of state, we can compare SEP tests with pulsars of different masses; in the weak field, we can also compare our result with tests based on the absence of dipole gravitational radiation. Dipole gravitational radiation can arise only if the centre of gravitational mass of a binary is not the same as the centre of inertial mass. Thus, dipole gravitational radiation implies an SEP violation, but relating such upper limits to direct SEP tests requires a specific theory. Within the quasi-Brans–Dicke framework, if we choose an equation of state then we can integrate the generalized Tolman–Oppenheimer–Volkoff equations[17] to compute both Δ for that neutron star and the degree to which it produces gravitational dipole radiation. Following ref. [22] we choose a very stiff equation of state ('.20' from ref. [25]; the maximum mass for a neutron star in this equation of state is 2.6 solar masses; stiffer equations of state lead to less-constraining limits from pulsar-based tests). In this family of theories, and with this equation of state, we evaluate constraints from our result, from existing wide binaries[50], from lunar laser ranging[4], from an SEP test measuring the location of the Solar System barycentre with MESSENGER[3], from the Cassini Shapiro delay measurement[20] and from gravitational-dipole-radiation upper limits on two pulsar–white dwarf systems[21,22]. The results are plotted in Fig. 3. All upper limits are at the 95% confidence level. The parameterized post-Newtonian parameters β and γ are proportional to α_0^2, so comparing the parameterized post-Newtonian values doubles the number of orders-of-magnitude difference between tests relative to comparing α_0. The constraint derived from the motion of PSR J0337+1715, in addition to being a direct strong-field test of the SEP, improves substantially on existing theory constraints for most values of β_0.

Because this family of theories supports spontaneous scalarization, comparing constraints becomes complicated when β_0 is less than about -4; the occurrence of spontaneous scalarization depends on both β_0 and the equation of state of neutron-star matter. Because spontaneous scalarization gives a pulsar a scalar coupling of order unity regardless of the weak-field behaviour of the theory, almost any theory constraint on a pulsar rules out the occurrence of spontaneous scalarization for pulsars with similar masses. Our constraint does not substantially assist in this process, because the mass of the pulsar in PSR J0337+1715 (1.4359 solar masses) is not very different from that in PSR 1738+0333. The pulsar mass ranges for which spontaneous scalarization remains possible[24] are around 1.6 solar masses and above 2 solar masses; ruling these out will require either (indirect) SEP tests on pulsars in these mass ranges or ruling out certain equations of state. These SEP tests may come about as a result of gravitational-wave observations of inspirals involving neutron stars of appropriate masses. Expected sensitivities suggest that even Advanced LIGO may not be able to carry out these tests to sufficient accuracy, but that its successors, Cosmic Explorer and the Einstein Telescope, may be able to rule out spontaneous scalarization if suitable inspirals are detected[24]. Precision timing of other pulsar binaries may also be able to rule out spontaneous scalarization if pulsars in suitable mass ranges are found. Studies of the neutron-star equation of state, such as those using NICER[51], may also rule out spontaneous scalarization. When spontaneous scalarization does not occur, constraints on alternative theories are determined almost entirely by the sensitivity of current tests; in this regime, as shown in Fig. 3, the constraint we derive from PSR J0337+1715 sets the current limits, and gravitational-wave observations even with the Cosmic Explorer and Einstein Telescope are not likely to improve on them[24]. On the contrary, the insensitivity of these telescopes to energy losses due to dipole gravitational-wave losses implies that these effects can be ignored when generating template gravitational-wave signatures for the Cosmic Explorer and Einstein Telescope when spontaneous scalarization does not occur. Because templates are generated for

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

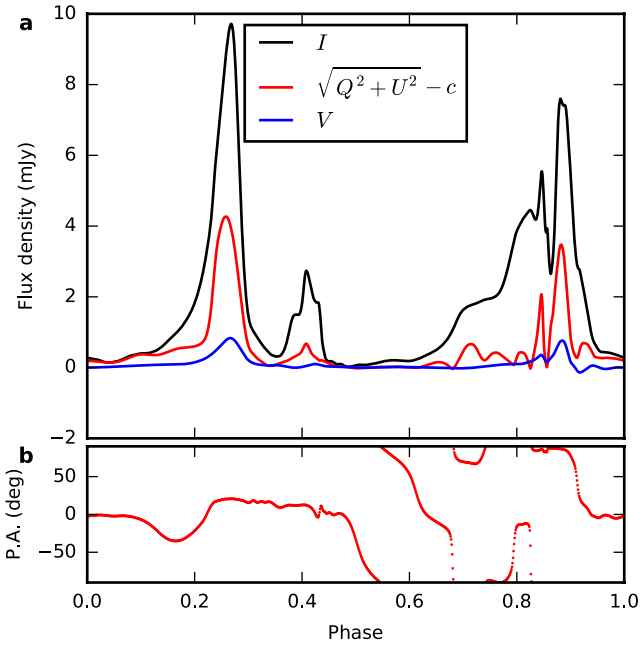
known neutron-star masses, only the masses for which spontaneous scalarization has not yet been ruled out need alternative-theory templates; and even these masses need to be considered only for the equations of state and values of β_0 that give rise to substantial scalarization.

Code availability. All general-purpose software packages that we used are open-source: PSRCHIVE (http://psrchive.sourceforge.net/), TEMPO (http://tempo.sourceforge.net/), TEMPO2 (http://www.atnf.csiro.au/research/pulsar/tempo2/), numdifftools (https://pypi.python.org/pypi/Numdifftools), sympy (http://www.sympy.org/), boost (http://www.boost.org/), galpy (https://github.com/jobovy/galpy) and GalDynPsr (https://github.com/pathakdhruv/GalDynPsr). The software written specifically for this project is available at https://doi.org/10.5281/zenodo.1241130.

Data availability. Data and ephemerides are available from the corresponding author on request. Pulse arrival time data, annotated with derivatives and residuals, and data for Fig. 3 and Extended Data Fig. 1 are available at https://doi.org/10.5281/zenodo.1241121. Source Data for Fig. 3 and Extended Data Fig. 1 are also available with the online version of the paper.

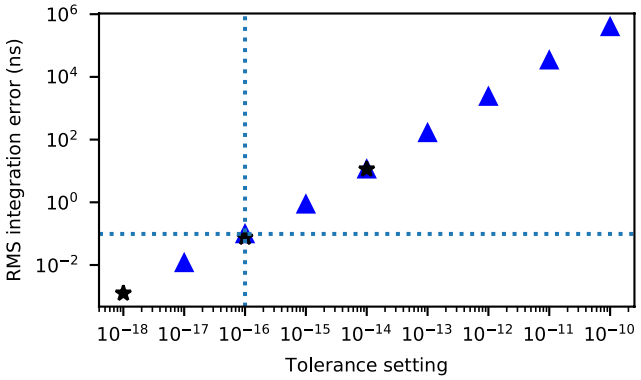
26. DuPlain, R. et al. Launching GUPPI: the Green Bank ultimate pulsar processing instrument. *Proc. SPIE* **7019**, 70191D (2008).
27. Karuppusamy, R., Stappers, B. & van Straten, W. PuMa-II: a wide band pulsar machine for the Westerbork Synthesis Radio Telescope. *Publ. Astron. Soc. Pacif.* **120**, 191–202 (2008).
28. Hankins, T. H. & Rickett, B. J. in *Methods in Computational Physics. Volume 14 – Radio Astronomy* (eds Alder, B. et al.) 55–129 (Academic Press, New York, 1975).
29. van Straten, W. Radio astronomical polarimetry and high-precision pulsar timing. *Astrophys. J.* **642**, 1004–1011 (2006).
30. Lambert, H. C. & Rickett, B. J. On the theory of pulse propagation and two-frequency field statistics in irregular interstellar plasmas. *Astrophys. J.* **517**, 299–317 (1999).
31. Archibald, A. M., Kondratiev, V. I., Hessels, J. W. T. & Stinebring, D. R. Millisecond pulsar scintillation studies with LOFAR: initial results. *Astrophys. J.* **790**, L22 (2014).
32. Bulirsch, R. & Stoer, J. Asymptotic upper and lower bounds for results of extrapolation methods. *Numer. Math.* **8**, 93–104 (1966).
33. Hobbs, G. B., Edwards, R. T. & Manchester, R. N. TEMPO2, a new pulsar-timing package – I. An overview. *Mon. Not. R. Astron. Soc.* **369**, 655–672 (2006).
34. Kopeikin, S. M. On possible implications of orbital parallaxes of wide orbit binary pulsars and their measurability. *Astrophys. J.* **439**, L5–L8 (1995).
35. Nordtvedt, K. Testing relativity with laser ranging to the Moon. *Phys. Rev.* **170**, 1186–1187 (1968).
36. Thirring, H. Über die Wirkung rotierender ferner Massen in der Einsteinschen Gravitationstheorie. *Phys. Z.* **19**, 33 (1918).
37. Einstein, A. *Über Gravitationswellen* 154–167 (Königlich-Preußischen Akademie der Wissenschaften, Berlin, 1918).
38. Russell, H. N. On the advance of periastron in eclipsing binaries. *Mon. Not. R. Astron. Soc.* **88**, 641–643 (1928).
39. Smarr, L. L. & Blandford, R. The binary pulsar: physical processes, possible companions, and evolutionary histories. *Astrophys. J.* **207**, 574–588 (1976).
40. Gusinskaia, N. V. et al. Conquering systematics in the timing of the pulsar triple system J0337+1715: towards a unique and robust test of the strong equivalence principle. *J. Phys. Conf. Ser.* **932**, 012003 (2017).
41. Prodan, S. & Murray, N. On the dynamics and tidal dissipation rate of the white dwarf in 4U 1820−30. *Astrophys. J.* **747**, 4 (2012).
42. Wex, N. in *Frontiers in Relativistic Celestial Mechanics. Volume 2: Applications and Experiments* (ed. Kopeikin, S. M.) 39–102 (De Gruyter, Berlin, 2014).
43. Luan, J. & Goldreich, P. Secular evolution of the pulsar triple system J0337+1715. *Astrophys. J.* **790**, 82 (2014).
44. Di Casola, E., Liberati, S. & Sonego, S. Nonequivalence of equivalence principles. *Am. J. Phys.* **83**, 39–46 (2015).
45. Damour, T. & Esposito-Farèse, G. Testing gravity to second post-Newtonian order: a field-theory approach. *Phys. Rev. D* **53**, 5541–5578 (1996).
46. Horbatsch, M. W. & Burgess, C. P. Model-independent comparisons of pulsar timings to scalar-tensor gravity. *Class. Quantum Gravity* **29**, 245004 (2012).
47. Bekenstein, J. D. Relativistic gravitation theory for the modified Newtonian dynamics paradigm. *Phys. Rev. D* **70**, 083509 (2004).
48. Shao, L. & Wex, N. Tests of gravitational symmetries with radio pulsars. *Sci. China Phys. Mech. Astron.* **59**, 699501 (2016).
49. Taylor, J. H., Wolszczan, A., Damour, T. & Weisberg, J. M. Experimental constraints on strong-field relativistic gravity. *Nature* **355**, 132–136 (1992).
50. Stairs, I. H. et al. Discovery of three wide-orbit binary pulsars: implications for binary evolution and equivalence principles. *Astrophys. J.* **632**, 1060–1068 (2005).
51. Arzoumanian, Z. et al. The neutron star interior composition explorer (NICER): mission definition. *Proc. SPIE* **9144**, 914420 (2014).
52. Irwin, A. W. & Fukushima, T. A numerical time ephemeris of the Earth. *Astron. Astrophys.* **348**, 642–652 (1999).
53. Lorimer, D. R. & Kramer, M. in *Handbook of Pulsar Astronomy* Ch. 8, Appendix 2 (Cambridge Univ. Press, Cambridge, 2004).
54. Shklovskii, I. S. Possible causes of the secular increase in pulsar periods. *Sov. Astron.* **13**, 562–565 (1970).
55. Pathak, D. & Bagchi, M. GalDynPsr: A package to estimate dynamical contributions in the rate of change of the period of radio pulsars. Preprint at https://arxiv.org/abs/1712.06590 (2017).
56. Bovy, J. galpy: a python library for galactic dynamics. *Astrophys. J. Suppl. Ser.* **216**, 29 (2015).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



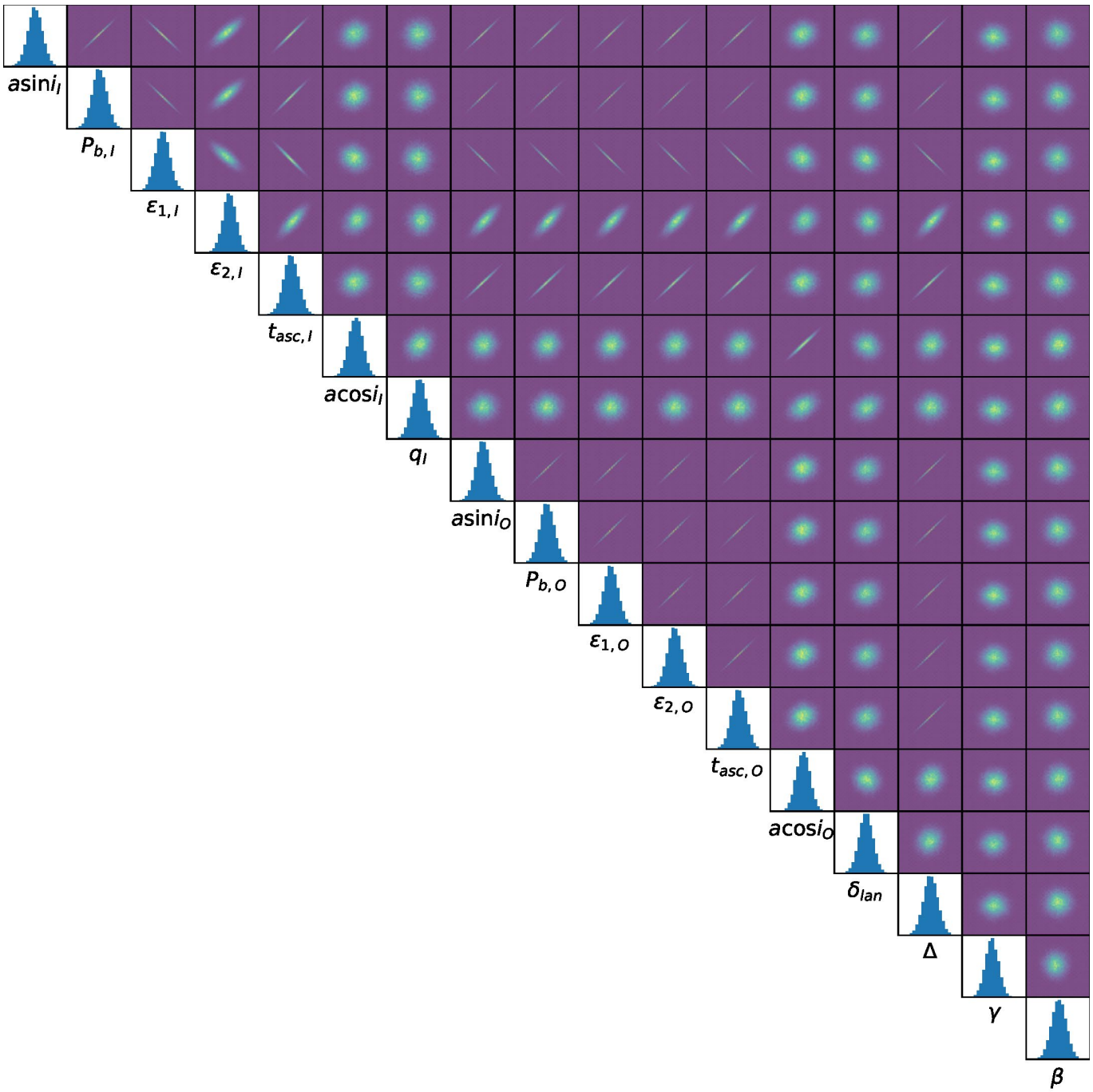
Extended Data Fig. 1 | Template pulse profile used for timing. This figure is based on the average 1,300–1,900-MHz profile from the GBT observation on MJD 56,412. The Stokes *IQUV* data have been smoothed by a wavelet-based algorithm (*psrsmooth*, PSRCHIVE). **a**, Total intensity (I) and linear (Q, U) and circular (V) polarization after correcting for Faraday rotation. An offset c has been subtracted; see below. **b**, Polarization angle (P.A.) at the centre frequency of the observation. The linear polarization (red) at some phases is responsible for almost half the flux density and its profile has complicated polarization structure. Offsets have been added to I and to $\sqrt{Q^2+U^2}$ to ensure that $I^2 \geq Q^2+U^2+V^2$.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



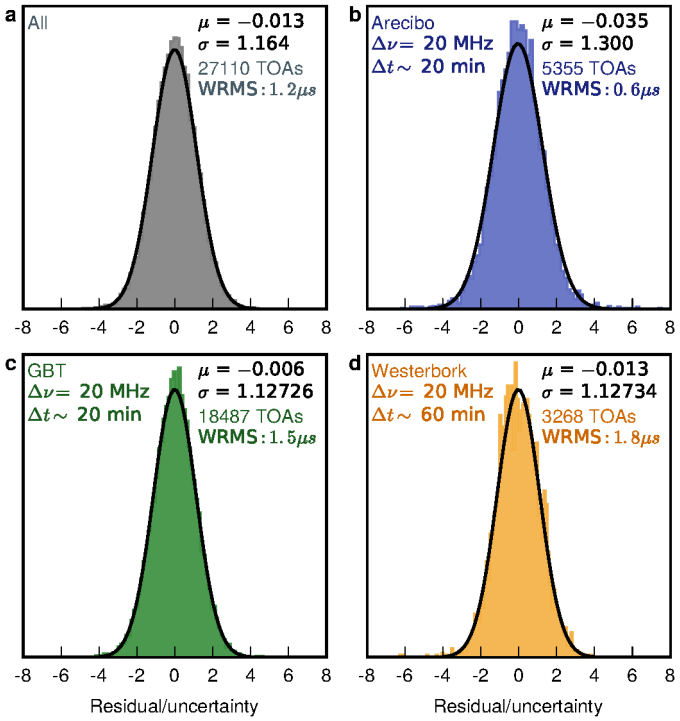
Extended Data Fig. 2 | Timing-model truncation error. The root-mean-square (RMS) arrival-time error caused by the finite time steps of the orbital integrator is shown as a function of the tolerance parameter. The vertical dotted line is the value used for all orbits in this work; the RMS error from truncation is below 0.1 ns. Blue triangles are calculations done in hardware 80-bit floating point; black stars are calculations done in software 128-bit floating point, which are much slower to compute. To estimate the errors in this plot, we compute a fiducial solution with 128-bit precision and a tolerance parameter of 10^{-22} and compare all other solutions to this one.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 3 | Covariances between parameters that affect the orbit. This plot does not include the parameters that are evaluated by linear least-squares fitting and marginalized out. Plots on the diagonal are single-parameter histograms; plots off the diagonal are pairwise two-dimensional histograms. See Extended Data Table 2 for parameter definitions.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 4 | Distribution of residuals divided by uncertainty, for each telescope. The standard deviation σ represents the factor by which the scatter of the post-fit residuals exceeds the claimed uncertainties on pulse arrival times; μ is the mean of the distribution. Each colour represents a different telescope. Only observations in the 1,400-MHz frequency band are shown here. Here $\Delta\nu$ and Δt are the bandwidth and time, respectively, over which the data are averaged to produce each pulse arrival time.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Fixed values and characteristics of the dataset

Parameter	Symbol	Value
Right ascension*†	RA	$03^h37^m43^s.82589$
Declination*†	Dec	$+17°15'14''.828$
Parallax†‡	π	0.770 milliarcseconds (mas)
Dispersion measure§	DM	$21.315933\,\mathrm{pc\,cm^{-3}}$
Solar system ephemeris		DE435
Time conversion ephemeris		IF99[36]
Time scale		TCB‖
Reference epoch		MJD 55920.0
Observation span		MJD 55956.7–57866.9
Number of TOAs		27110
Root-mean-squared residual (weighted)		$1.2\,\mu$s

The time conversion ephemeris is from ref. [52].
*We use the same position as in the discovery paper[8].
†Although these parameters were held fixed, we fit for offsets from them; see Extended Data Table 2.
‡We use the distance estimate, based on white-dwarf modelling, given in the discovery paper[8].
§We determine the dispersion measure from a global fit that assumes no dependence of the profile on frequency.
‖TCB is barycentric coordinate time, a timescale that runs at a slightly different rate (faster by 1.550505×10^{-8}) than terrestrial clocks because it has been corrected for the gravitational time dilation due to the Solar System potential.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 2 | Fitted values

Parameter	Symbol	Value
Pulsar spin parameters		
Pulsar spin frequency*†	f	$365.953363080(4)$ Hz
Pulsar spin frequency derivative	$\dot{f}$	$-2.355208(3) \times 10^{-15}$ Hz s^{-1}
Astrometric parameters‡		
Right ascension offset	ΔRA	$10.4(2)$ mas
Declination offset	ΔDEC	$-22.8(10)$ mas
Proper motion in right ascension	μ_{RA}	$4.51(6)$ mas/yr
Proper motion in declination	μ_{DEC}	$2.2(2)$ mas/yr
Parallax offset	$\Delta\pi$	$-7.5(12) \times 10^{-2}$ mas
Inner Keplerian parameters for pulsar orbit		
Semi-major axis projected along line of sight	$(a \sin i)_I$	$1.2175252(2)$ lt-s
Orbital period	$P_{b,I}$	$1.6293932(6)$ d
Eccentricity parameter§ $(e \sin \omega)$	$\epsilon_{1,I}$	$6.8833(20) \times 10^{-4}$
Eccentricity parameter $(e \cos \omega)$	$\epsilon_{2,I}$	$-9.1401(17) \times 10^{-5}$
Time of ascending node	$t_{\mathrm{asc},I}$	MJD $55920.40771662(6)$
Outer Keplerian parameters for centre of mass of inner binary		
Semi-major axis projected along line of sight	$(a \sin i)_O$	$74.672629(13)$ lt-s
Orbital period	$P_{b,O}$	$327.25685(11)$ d
Eccentricity parameter $(e \sin \omega)$	$\epsilon_{1,O}$	$3.518595(5) \times 10^{-2}$
Eccentricity parameter $(e \cos \omega)$	$\epsilon_{2,O}$	$-3.46313(17) \times 10^{-3}$
Time of ascending node	$t_{\mathrm{asc},O}$	MJD $56233.93512(11)$
Orbital interaction parameters		
Semi-major axis projected in plane of sky	$(a \cos i)_I$	$1.48950(19)$ lt-s
Semi-major axis projected in plane of sky	$(a \cos i)_O$	$91.358(12)$ lt-s
Ratio of inner comp. mass to PSR mass	$q_I = m_{cI}/m_p$	$0.137405(4)$
Difference in longs. of asc. nodes‖	δ_{lan}	$1.2(4) \times 10^{-4}$ °
GR violation parameters		
PPN nonlinearity-of-gravity parameter¶	$\beta_{\mathrm{PPN}} - 1$	$0(3) \times 10^{-3}$
PPN spacetime curvature parameter	$\gamma_{\mathrm{PPN}} - 1$	$0(2) \times 10^{-5}$
SEP violation parameter#	Δ	$(-1.1 \pm 0.7) \times 10^{-6}$

*Values in parentheses represent 1σ errors in the last decimal place(s), as determined by our MCMC fitting.

†The proper timescale of the pulsar is slowed due to both gravitational time dilation and the transverse Doppler effect, in a way that varies as it moves around its orbit. Counter to the usual practice in pulsar timing, we do not correct this timescale so its average rate matches that on Earth.

‡To avoid having incorrect astrometry affect our SEP test, we use derivatives to fit for offsets between fixed astrometric parameters and astrometry as determined from timing. Because we have not analysed systematics affecting these parameters carefully we do not recommend further use of these timing-derived astrometric parameters.

§The Laplace–Lagrange parameters[53] ε_1 and ε_2 provide a parameterization of the eccentricity (e) and longitude of periastron (Ω) of an orbit that avoids a coordinate singularity at zero eccentricity; the pair $(\varepsilon_2, \varepsilon_1)$ forms a vector in the plane of the orbit called the eccentricity vector.

‖For a single orbit, the ascending node is the place where the pulsar passes through the plane of the sky moving away from us; the longitude of the ascending node specifies the orientation of the orbit on the sky. This is not measurable with the data that we have, but the difference between the longitudes of the ascending nodes of the two orbits is measurable through orbital interactions.

¶The parameterized post-Newtonian parameters β and γ are not substantially constrained by our observations and their values and uncertainties are consistent with the priors that we obtained from Solar System experiments.

#The SEP-violation parameter Δ is the fractional difference in acceleration between the pulsar and the inner white dwarf; because it is the focus of this paper we take additional steps to estimate the effect of systematics on it.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 3 | Inferred values

Parameter	Symbol	Value
Pulsar properties		
Pulsar period	P	$2.73258863256(3)$ ms
Pulsar period derivative	$\dot{P}$	$1.758643(2) \times 10^{-20}$
Corrected pulsar period derivative*	$\dot{P}$	$1.7293(16) \times 10^{-20}$
Inferred surface dipole magnetic field†	B	2.2×10^{8} G
Spin-down power	$\dot{E}$	3.4×10^{34} erg s^{-1}
Characteristic age	τ	2.5×10^{9} y
Orbital geometry		
Pulsar semi-major axis (inner)	a_I	$1.92379(14)$ lt-s
Eccentricity (inner)	e_I	$6.9437(20) \times 10^{-4}$
Longitude of periastron (inner)	ω_I	$97.5638(14)$ °
Pulsar semi-major axis (outer)	a_O	$117.992(9)$ lt-s
Eccentricity (outer)	e_O	$3.535596(3) \times 10^{-2}$
Longitude of periastron (outer)	ω_O	$95.6212(3)$ °
Inclination of invariant plane‡	i	$39.262(4)$ °
Inclination of inner orbit	i_I	$39.263(4)$ °
Angle between orbital planes	δ_i	$1.4(6) \times 10^{-3}$ °
Angle between eccentricity vectors	$\delta_\omega \sim \omega_O - \omega_I$	$-1.9427(16)$ °
Relativistic periastron advance (inner)	$\dot{\omega}_I$	$0.122293(18)$ °/year
Relativistic periastron advance (outer)	$\dot{\omega}_O$	$2.0636(3) \times 10^{-5}$ °/year
Masses§		
Pulsar mass	m_p	$1.4359(3)\ M_\odot$
Inner companion mass	m_{cI}	$0.19730(4)\ M_\odot$
Outer companion mass	m_{cO}	$0.40962(9)\ M_\odot$

*This period derivative includes corrections for the Shklovskii effect[54] (using the unreliable astrometry computed here) and acceleration in the Galactic potential[55,56].
†We use the standard formulae[53] for computing B, $\dot{E}$ and τ; in particular, we assume a pulsar mass of 1.4 solar masses and a moment of inertia of 10^{45} g cm^2.
‡The invariant plane is the plane perpendicular to the total (orbital) angular momentum of the triple system.
§Masses are as measured by a distant observer; corrections for special relativity are at the 10^{-8} fractional level and are therefore irrelevant.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

https://doi.org/10.1038/s41586-018-0250-8

Multifunctional ferrofluid-infused surfaces with reconfigurable multiscale topography

Wendong Wang[1,2,3,13], Jaakko V. I. Timonen[1,4,13], Andreas Carlson[1,2,5], Dirk-Michael Drotlef[3], Cathy T. Zhang[1,2], Stefan Kolle[1], Alison Grinthal[1], Tak-Sing Wong[1,2,8], Benjamin Hatton[1,2,9], Sung Hoon Kang[1,2,10], Stephen Kennedy[1,2,11], Joshua Chi[1,6,12], Robert Thomas Blough[2], Metin Sitti[3], L. Mahadevan[1,2,7] & Joanna Aizenberg[1,2,6,7]*

Developing adaptive materials with geometries that change in response to external stimuli provides fundamental insights into the links between the physical forces involved and the resultant morphologies and creates a foundation for technologically relevant dynamic systems[1,2]. In particular, reconfigurable surface topography as a means to control interfacial properties[3] has recently been explored using responsive gels[4], shape-memory polymers[5], liquid crystals[6–8] and hybrid composites[9–14], including magnetically active slippery surfaces[12–14]. However, these designs exhibit a limited range of topographical changes and thus a restricted scope of function. Here we introduce a hierarchical magneto-responsive composite surface, made by infiltrating a ferrofluid into a microstructured matrix (termed ferrofluid-containing liquid-infused porous surfaces, or FLIPS). We demonstrate various topographical reconfigurations at multiple length scales and a broad range of associated emergent behaviours. An applied magnetic-field gradient induces the movement of magnetic nanoparticles suspended in the ferrofluid, which leads to microscale flow of the ferrofluid first above and then within the microstructured surface. This redistribution changes the initially smooth surface of the ferrofluid (which is immobilized by the porous matrix through capillary forces) into various multiscale hierarchical topographies shaped by the size, arrangement and orientation of the confining microstructures in the magnetic field. We analyse the spatial and temporal dynamics of these reconfigurations theoretically and experimentally as a function of the balance between capillary and magnetic pressures[15–19] and of the geometric anisotropy of the FLIPS system. Several interesting functions at three different length scales are demonstrated: self-assembly of colloidal particles at the micrometre scale; regulated flow of liquid droplets at the millimetre scale; and switchable adhesion and friction, liquid pumping and removal of biofilms at the centimetre scale. We envision that FLIPS could be used as part of integrated control systems for the manipulation and transport of matter, thermal management, microfluidics and fouling-release materials.

Whereas the magnetic field-induced reconfiguration of a ferrofluid to form macroscopic protuberances on a flat surface is well known[15,16], the behaviour of a ferrofluid in a microstructured confinement might elicit a range of otherwise unachievable multiscale topographical responses, enabled by the capillary pressure within the porous substrate (Fig. 1a). To explore this concept, we used various microstructured substrates (see Table 1) infiltrated with fluorocarbon- or silicone-oil-based ferrofluids. A small permanent magnet was placed under the samples and the resulting spatiotemporal changes in surface topography were visualized using an angled illumination technique. The thickness of the ferrofluid overlayer was measured using a force probe (see Extended Data Fig. 1) and the evolution of the profile of the ferrofluid–air interface was measured using a laser scanning microscope (Extended Data Fig. 2a, b).

The non-uniform magnetic field created by the magnet initiates three sequential and interrelated processes to generate dynamic multiscale topographies (Fig. 1b). In the first step, the magnetic field causes a withdrawal of the initially flat ferrofluid overlayer above the microstructures. Withdrawal leads to the formation of a macroscopic protuberance with lateral size comparable to that of the magnet (about 1–20 mm). In the second step, the ferrofluid that remains trapped in the microstructured matrix through capillary force is pulled out of the pores by a magnetic pressure $|p_m| \approx \mu_0 M_s H_0$ (Extended Data Fig. 3a–c), where μ_0 (in units of N A^{-2}) is the vacuum permeability, M_s (A m^{-1}) is the saturation magnetization of the ferrofluid and H_0 (A m^{-1}) the strength of the applied magnetic field[15]. The magnetic pressure is counteracted by the capillary pressure $p_\gamma \approx 2\gamma/d_y$, where γ is the surface tension of the ferrofluid–air interface and $d_y/2$ is half the width of the channels and the largest characteristic radius of the porous matrix. If $|p_m| \leq p_\gamma$, then ferrofluid will remain trapped in the pores (Extended Data Fig. 3e–g). If $|p_m| > p_\gamma$, such as for the case shown in Fig. 1b and Extended Data Fig. 3d, where $|p_m| \approx 10^4$ Pa and $p_\gamma \approx 10^3$ Pa, then ferrofluid will be extracted from the pores, leading to the appearance of the conformally coated micro-topographical region (1) (Fig. 1a). In the third step, the micro-topographical area expands outwards through porous-capillary flow, while the macroscopic protuberance continues to grow through the accumulation of ferrofluid. The initial extraction of ferrofluid from the areas around the magnet deforms the ferrofluid–air interface, producing a capillary pressure even in areas far from the magnet where the magnetic pressure is smaller than the capillary pressure. The interface has a gradually increasing height along the channel, from its minimum near the magnet to its maximum height h_0 far away from the magnet (Fig. 1c, Extended Data Fig. 2a). The combination of the magnetic and capillary pressures makes the flow follow the micro-topography, even in highly complex channel geometries such as the spiral shape of pattern 6 (Table 1), where the flow makes turns along a curved path (see Extended Data Fig. 2e).

The size of the micro-topographical area, characterized by its length along the micro-channels L_x (shown in state (iv) of Fig. 1b), was observed to scale with time as $L_x \propto t^{0.35-0.5}$ for different patterns and overlayer thicknesses (Fig. 1d, Extended Data Fig. 2c, d). By balancing the rate of change of work done by the capillary force with the viscous dissipation, we obtain

$$\gamma \dot{L}_x d_y \approx \gamma d_y U \approx \int \eta \left(\frac{\partial U}{\partial z}\right)^2 \mathrm{d}V \approx \eta \left(\frac{U}{h_0}\right)^2 h_0 L_x d_y \qquad (1)$$

[1]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. [2]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA, USA. [3]Max Planck Institute for Intelligent Systems, Stuttgart, Germany. [4]Department of Applied Physics, Aalto University School of Science, Espoo, Finland. [5]Department of Mathematics, Mechanics Division, University of Oslo, Oslo, Norway. [6]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [7]Kavli Institute for Bionano Science and Technology, Harvard University, Cambridge, MA, USA. [8]Present address: Department of Mechanical and Nuclear Engineering and the Materials Research Institute, The Pennsylvania State University, University Park, PA, USA. [9]Present address: Materials Science and Engineering, University of Toronto, Toronto, Ontario, Canada. [10]Present address: Department of Mechanical Engineering, Hopkins Extreme Materials Institute, Johns Hopkins University, Baltimore, MD, USA. [11]Present address: Department of Biomedical and Chemical Engineering, University of Rhode Island, Kingston, RI, USA. [12]Present address: Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, USA. [13]These authors contributed equally: Wendong Wang, Jaakko V. I. Timonen. *e-mail: jaiz@seas.harvard.edu

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

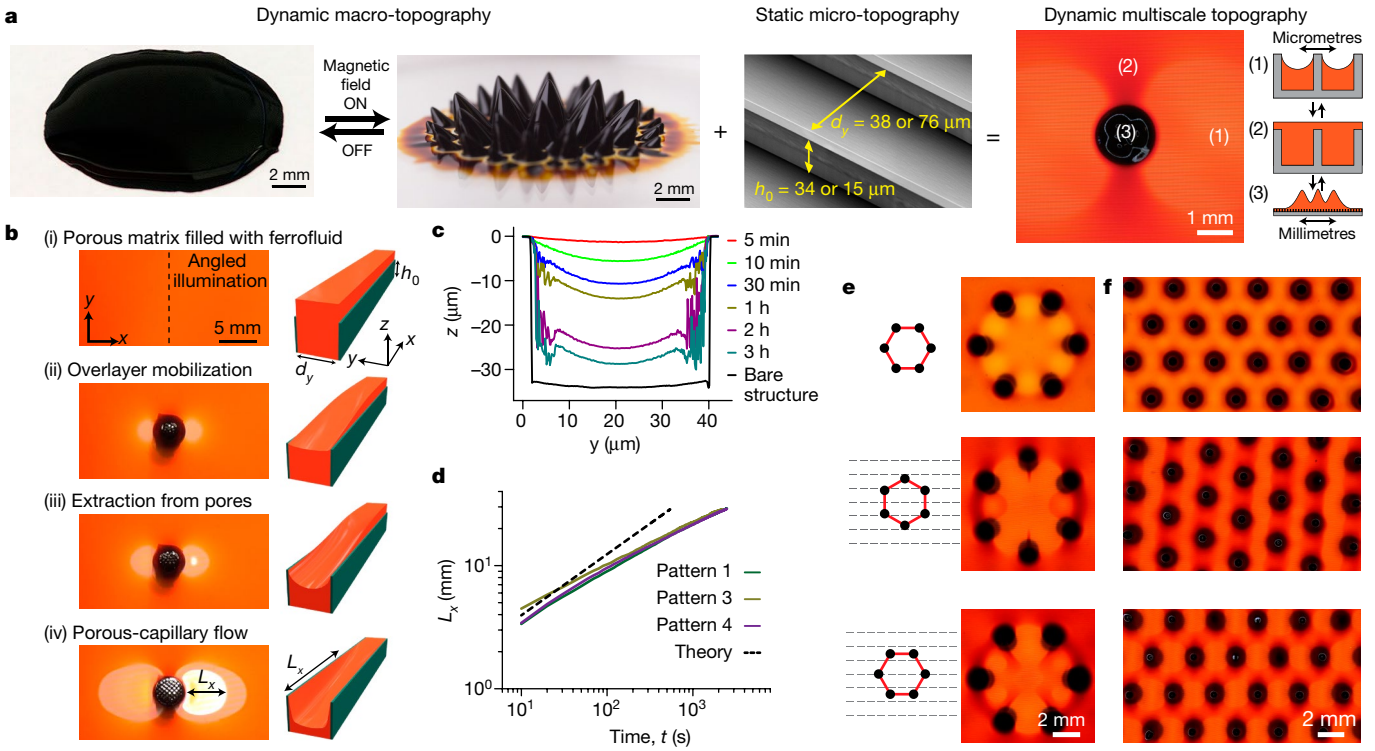


Fig. 1 | Dynamic multiscale topography of ferrofluid-containing liquid-infused porous surfaces (FLIPS). a, A diagram showing the concept of FLIPS. Left, two topographical states of a ferrofluid spreading on an unstructured surface, depicting the transformation from the flat interface to macroscale protuberances in response to an external magnetic field. Middle, scanning electron micrograph of the static micro-topography of exemplary microchannel substrates 1–3. Right, top-view photo of FLIPS under magnetic field, showing that the ferrofluid confined within the microstructured solid experiences area-selective topographical reconfigurations at multiple length scales: region (1) exhibits micro-topography shaped by the structured substrate; region (2) exhibits a flat surface; and region (3) exhibits the macro-topographical protuberance. **b**, Transport processes involved in the formation of macro- and micro-topographical features: the left column shows a series of representative experimental photos in top view, captured using the angled illumination technique; the right column shows the corresponding schematics, depicting the deformation of the ferrofluid–air interface in a microchannel. **c**, Evolution of cross-sectional profiles of the ferrofluid–air interface at a fixed distance (about 1.5 cm away from the magnet) over time, measured using a laser scanning microscope. **d**, Logarithmic plots of experimentally measured L_x versus time t for patterns 1, 3 and 4 from Table 1. The dashed black line is plotted using equation (2), with $\gamma = 17\,\mathrm{mN\,m^{-1}}$, $\eta = 0.367\,\mathrm{Pa\,s}$ and $h_0 = 34\,\mu\mathrm{m}$. The overlayer thicknesses are ~10–20 μm. **e**, **f**, Snapshots of multiscale topographical response to a hexagonal pattern of six (**e**) or more (**f**) magnets (diameter, 1.6 mm; magnet spacing, 3.2 mm). The schematics on the left in **e** show the relative orientation of magnets with respect to microstructures. In **e** and **f**, the top row shows the case without a microstructured substrate; the bottom two rows show cases with the microstructured pattern 4, with different orientations of the hexagonal pattern.

where the speed of the advancing front is $U \approx \dot{L}_x$ and $\dot{L}_x = \mathrm{d}L_x/\mathrm{d}t$. Rearranging the terms and integrating with respect to time t gives a scaling relation for L_x:

$$L_x(t) \approx \left(\frac{\gamma h_0}{\eta}\right)^{1/2} t^{1/2} \qquad (2)$$

where γ is the surface tension, η is the dynamic viscosity and h_0 is the height of the microstructure. The magnetic pressure $|p_\mathrm{m}|$ generates a force on the ferrofluid, and an alternative derivation to the scaling relationship in equation (2) is obtained by replacing the left-hand side of equation (1) by the rate of change of work done by the magnet on the ferrofluid ($|p_\mathrm{m}|\dot{L}_x d_y h_0$) and integrating with respect to time:

$$L_x(t) \approx \left(\frac{|p_\mathrm{m}|h_0^2}{\eta}\right)^{1/2} t^{1/2} \qquad (3)$$

Both derivations give the same power-law dependence with respect to time $L_x \propto t^{1/2}$, differing only in their pre-factors. Despite their simplicity, these scaling relations provide a reasonable order-of-magnitude approximation of the dynamics revealed in the experimental data. They also capture the dependence of L_x on h_0 and the independence of L_x from the spacing between micro-plates along the x direction d_x (see Extended Data Fig. 1a) and d_y. The discrepancy between the experiments and these scaling laws may be due to the three-dimensional shape of the microchannels, the complex shape of the ferrofluid–air interface and the dependence of L_x on the thickness of the ferrofluid overlayer (see Extended Data Fig. 2c), which the current models do not fully capture.

The asymmetry of the microchannels used in the demonstration above induces preferential asymmetric extraction of the ferrofluid from the channels along the x direction and the appearance of the characteristic dumbbell-shaped micro-topographical area (Supplementary Video 1). Additional length scales can be introduced by using patterned or structured magnetic fields applied to geometrically anisotropic FLIPS. Such fields can be created by organizing multiple permanent magnets into an array. For example, a hexagonal cluster of six magnets acting on a FLIPS with a channel-like array of microplates introduces one more symmetry element to the system, leading to more complicated flow patterns that reflect the relative orientation of the anisotropic microstructures in the patterned magnetic field (Fig. 1e). Because the field source is no longer axisymmetric, the orientation of the field source (six-fold rotational symmetry) with respect to the microplates (two-fold rotational symmetry) allows for one more degree of control over the topographical response (Fig. 1e, Supplementary Video 1), with the opportunity to expand the combination of symmetry elements in FLIPS to any combination of magnet assembly (for example, 'infinite' arrays of magnets; see Fig. 1f) and structured surface.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Table 1 | Structured surfaces used in this study

Pattern	Type	Dimensions (μm)
1	Array of microchannels	$d_x=0$, $d_y=38$, $h_0=34$
2	Array of microchannels	$d_x=0$, $d_y=38$, $h_0=15$
3	Array of microchannels	$d_x=0$, $d_y=76$, $h_0=34$
4	Array of microplates	$d_x=5$, $d_y=38$, $h_0=30$
5	Array of microposts	$d_x=1.4$, $d_y=1.4$, $h_0=10$
6	Spirally shaped channels	$d_x=0$, $d_y=38$, $h_0=15$
7	Microporous membrane	Average pore size = 1
8	Microporous membrane	Average pore size = 10
9	Microporous membrane	Average pore size = 20
10	Microporous tubing	Average pore size = 5–60

The FLIPS concept introduced here offers substantial versatility in designing dynamic surfaces with multiscale topographical responses. In the examples shown, we demonstrate that the specific topographical patterns can be finely tuned by controlling: (i) the properties of the ferrofluid (such as the type of magnetic particle, the concentration of magnetic particles, or the type and viscosity of the carrier fluid); (ii) the geometry of the microstructured substrate; (iii) the strength and pattern of the magnetic field; and (iv) the distance of the FLIPS from the magnets and their relative orientation. The resulting spatial and temporal dynamics of topographical reconfiguration enable numerous functions at multiple length scales.

At the micrometre scale, FLIPS provide a way of manipulating colloidal matter on two-dimensional interfaces (Fig. 2). Depending on the state of the dynamic topography, colloidal particles can form a disordered two-dimensional gas-like state (on flat topography; Fig. 2a) or organize into structures such as chains (on micro-topographical regions; Fig. 2b). Subsequently, FLIPS allow controlled transport of non-magnetic colloids when a horizontal body force on the ferrofluid is exerted by a lateral magnetic-field gradient (Fig. 2c, d). The transport mechanism is unique, because it does not correspond to either positive or negative magnetophoresis of magnetic colloidal matter, which have been studied extensively before[20]. Instead, the force on the non-magnetic particle is created by a hydrodynamic coupling of the ferrofluid flow under the field gradient to the motion of the aqueous phase on top of the ferrofluid (Fig. 2e). Under typical experimental conditions using small permanent magnets as field sources, the speed of the colloids is of the order of a few micrometres per second and controllable with the magnetic field (Fig. 2f, Supplementary Video 2).

Multiscale topography can also be used to assemble colloidal matter into otherwise inaccessible hierarchical structures. For example, in the case of ferrofluid-coated unstructured surfaces, a hexagonal, soft, ferromagnetic nickel grid leads to the formation of a hexagonally varying, periodic ferrofluid pattern on which colloidal particles assemble into close-packed clusters (Fig. 2g); in the case of FLIPS, a multiscale topographical response is created, and colloidal matter additionally organizes into short line segments as dictated by the two symmetries (Fig. 2h).

At the millimetre scale, FLIPS can be harnessed for controlling the motion, clustering and interaction of liquid droplets. We use gravity

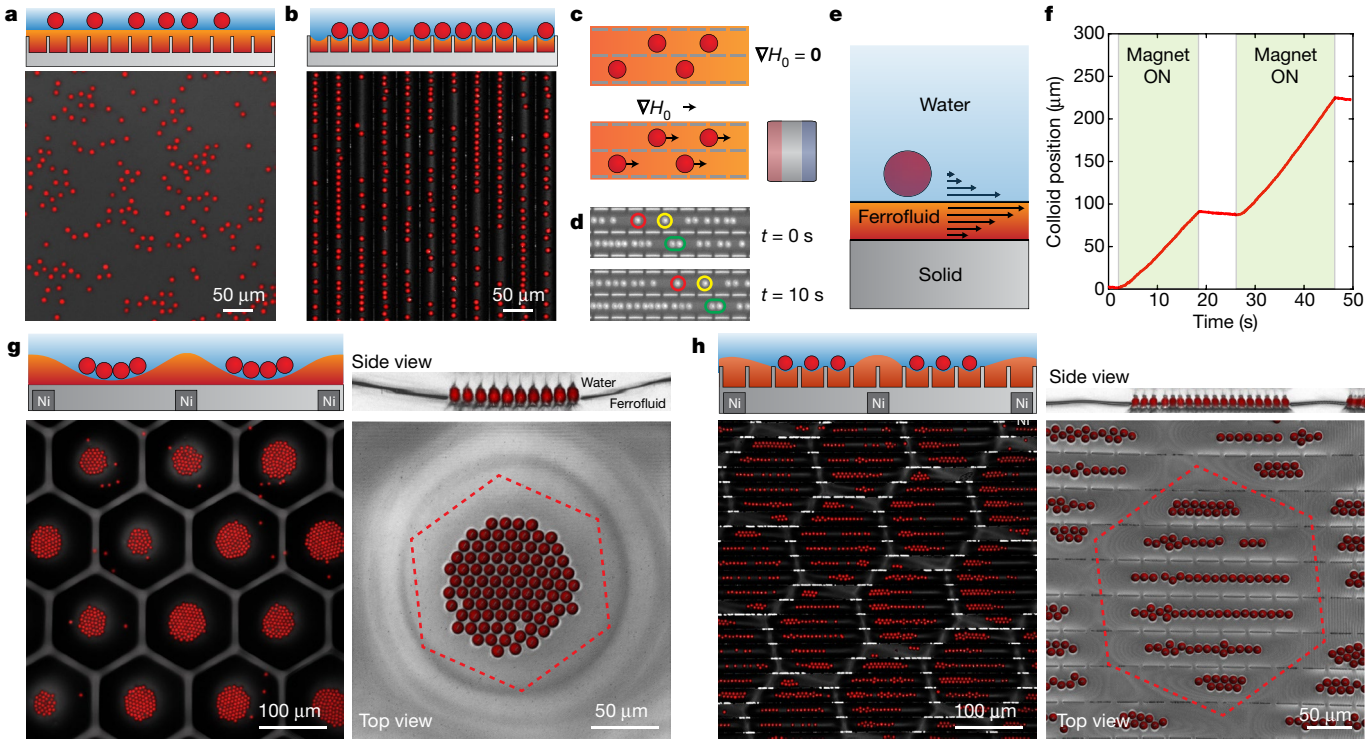


Fig. 2 | Manipulation of non-magnetic colloidal particles on FLIPS (application of FLIPS at micrometre scale). a, **b**, Schematics (top) and confocal fluorescence images (bottom) of 10-μm-diameter melamine colloidal particles floating on the flat surface of FLIPS in the absence of a magnetic field (**a**) and confined in micro-topography of FLIPS when the ferrofluid has been locally depleted from the microstructures using a magnet (**b**). **c**, Schematics showing colloidal particles staying stationary in the absence of a magnetic-field gradient ($\nabla H_0 = \mathbf{0}$; top) and moving in the presence of a magnetic-field gradient that is created by a magnet on the right (direction of ∇H_0 indicated by the arrow; bottom). **d**, Optical images showing the transport of colloidal particles along the micro-topographical region. Four particles are labelled to depict their movement within a 10s interval. **e**, Schematic illustrating the mechanism behind the transport of non-magnetic colloidal particles in a magnetic field: the transport of ferrofluid induces a flow of the water near the ferrofluid–water interface, leading to the transport of the colloidal particles. **f**, Plot of relative colloidal positions along the channels as a function of time; the green shaded areas indicate the times when the magnet field is turned on. **g**, **h**, Schematics (top left) and confocal fluorescence (side views on the top right and top views at the bottom) images showing the confinement of colloidal particles by the macro-topographical response of a thin layer of ferrofluid alone (without micro-topography; **g**) and the confinement of colloidal particles by the macro- and micro-topographical response of FLIPS (**h**). A hexagonal nickel grid embedded in FLIPS is used to shape a nearly uniform external magnetic field into a hexagonally varying field-intensity pattern. Schematics are not shown to scale.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

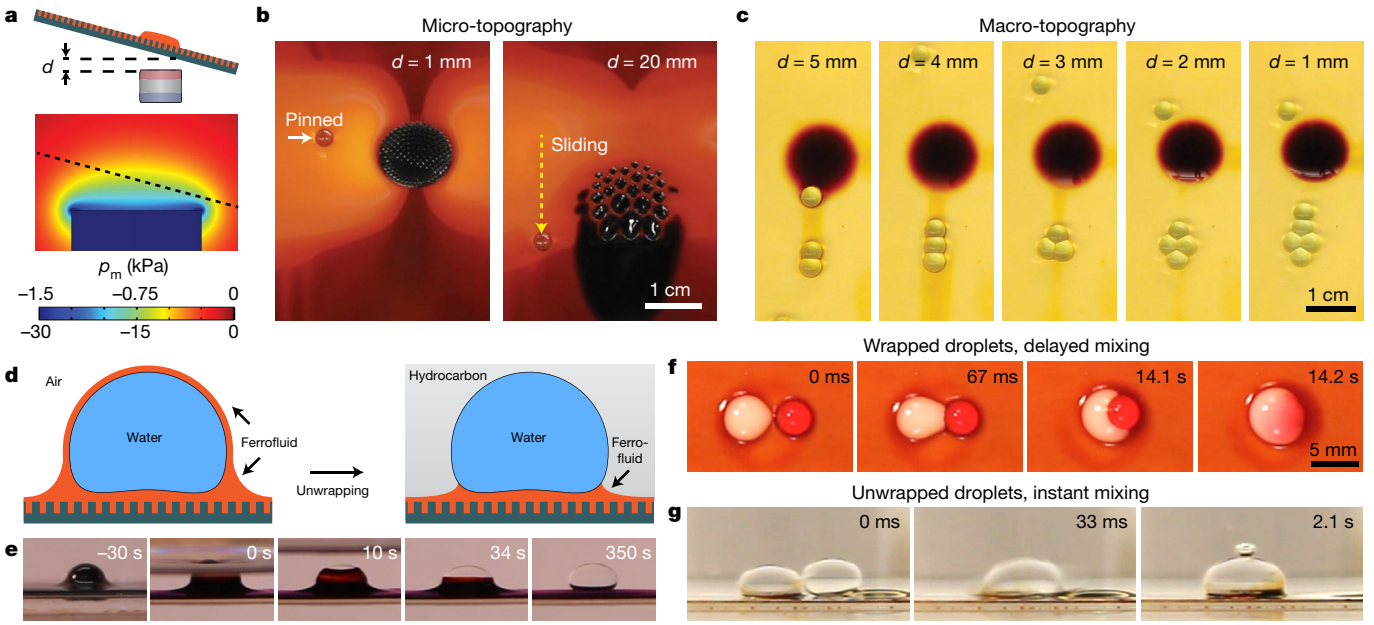


Fig. 3 | Control of droplet flow and droplet manipulation (application of FLIPS at millimetre scale). a, Top, schematic showing the configuration of the experimental set-up in **b** and **c**. The FLIPS is tilted and placed at a distance *d* above the magnet. Bottom, simulated magnetic pressure distribution (p_m) around the magnet. The upper and lower scales of the colour scale correspond to **c** and **b**, respectively. **b**, Two photographs showing a water droplet pinned on the micro-topographical area and its subsequent release after the magnet is lowered to allow ferrofluid to flow back and submerge the micro-topography. **c**, Five photographs showing different clustering behaviours of 15-μl water droplets on the macro-topographical feature at five different distances *d*. This FLIPS used diluted ferrofluid and hence did not have a micro-topographical area. **d**, Schematic showing the unwrapping of a thin ferrofluid layer around a water droplet by changing the surrounding medium from air to a hydrocarbon liquid. **e**, Five photographs showing the unwrapping of a thin layer of ferrofluid around a water droplet after the addition of dodecane. **f**, Four photographs showing the delayed mixing of two liquid droplets in the presence of wrapping layers. The white and red droplets are the suspension of 10-μm-diameter polystyrene colloids in ethanol–water mixture and in water dyed with rhodamine B, respectively. They were brought together by a stationary alternating-current electromagnet at the centre beneath the FLIPS. **g**, Three photographs showing instant mixing of two droplets in the absence of wrapping layers. The left droplet is an aqueous solution of sodium bicarbonate and the right droplet is an aqueous solution of 2 M hydrochloric acid. The right image shows the CO_2 bubble formed after the coalescence of the droplets.

to drive the flow of droplets on a tilted FLIPS, and tune the multiscale topographical response by adjusting the magnetic-field strength and gradients by varying the distance *d* between the magnet and FLIPS (Fig. 3a). On the one hand, the reversible appearance of micro-topography creates switchable slippery surfaces[21] that pin the droplets at micro-topographical regions or release the droplets when the magnetic-field strength is reduced to remove the micro-topography (Fig. 3b, Supplementary Video 3). On the other hand, the macro-topographical protuberance can be used to assemble droplets into well-defined clusters (Fig. 3c). To demonstrate this behaviour, we note that when a droplet slides near the protuberance it is pinned to it by capillary and magnetic forces. Subsequent droplets will enter the trap and combine with the previously pinned droplets to form clusters, until their collective gravity overcomes their attraction to the protuberance. They are then released as doublets, triplets, quadruplets or quintuplets, depending on the adhesion that is controlled by the distance between the magnet and the FLIPS (Supplementary Video 4). If the droplets contain polymerizable moieties, then these clusters can be solidified into distinct assemblies.

Another droplet-manipulation strategy involves control over the formation of the ferrofluid wrapping layer around the droplets. For example, by replacing air with an alternative liquid medium that is immiscible with the ferrofluid or the droplets, such as a hydrocarbon, the non-transparent ferrofluid wrapping layer around the water droplets can be removed (Fig. 3e, Supplementary Video 5). This unwrapping can be used to accelerate the coalescence of droplets: compared to the wrapped droplets, for which the mixing is delayed by more than 10 s (Fig. 3f), the unwrapped droplets mix instantaneously (Fig. 3g, Supplementary Video 6), enabling rapid initiation of chemical reactions. Although existing theory suggests that the formation of a wrapping layer is due to a positive spreading coefficient[12,22], we found that long-range van der Waals interaction is likely to be the decisive force in determining the presence or absence of the wrapping layers; see detailed discussion in Methods section 'Additional notes on the formation of the wrapping layer', Extended Data Tables 1–4 and Extended Data Figs. 4 and 5[23–26].

Finally, we illustrate the versatility of FLIPS at the centimetre scale by demonstrating switchable adhesion and friction, liquid pumping and biofilm removal. We measure the adhesion between FLIPS and other surfaces by first pressing the test surface against the FLIPS and then lifting them. With the magnetic field on, deforming the macroscopic protuberances and driving the ferrofluid back into the porous matrix requires external work. Conversely, from an energetic viewpoint, the formation of the macroscopic protuberance is favoured and can spontaneously open a gap between the FLIPS and the test surface, which enables switchable adhesion (Fig. 4a, Supplementary Video 7, Extended Data Fig. 6a–c) and friction (Supplementary Video 8). Notably, in contrast to traditional ferrofluid-coated unstructured surfaces, FLIPS display area-specific adhesion and friction: regions with flat ferrofluid, with macrostructured protuberances, and with micro-topography, have characteristically different values (Extended Data Fig. 6d–h). Furthermore, the anisotropic geometry and arrangement of the microstructures introduce directionality to friction that can be controlled magnetically (Extended Data Fig. 6i) to create tunable anisotropic-friction materials.

By coupling the ability of the magnets to extract ferrofluid from microporous structures and the movement of the magnets, we demonstrate pumping liquids at the centimetre scale (Fig. 4b, Supplementary Video 9). We infuse a porous polytetrafluoroethylene (PTFE) tube with ferrofluid to form a FLIPS pipe. A stepper motor moves five pairs of magnets in a circular motion. The ferrofluid between each pair of magnets moves as the magnets rotate and pushes the test liquid along the tube. The porosity of the PTFE tube wall allows the ferrofluid to

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

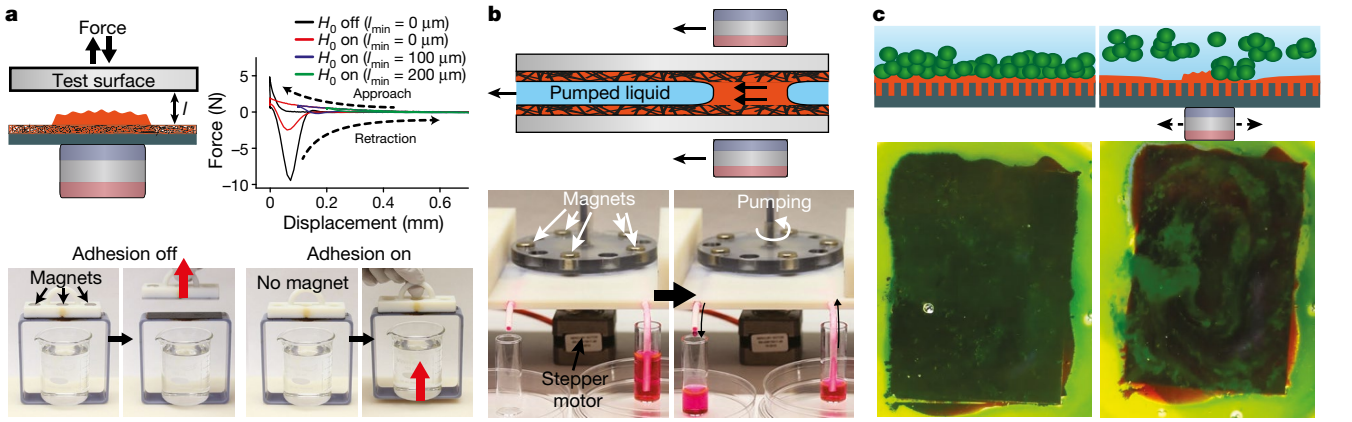


Fig. 4 | Adhesion, pumping and biofilm removal (application of FLIPS at centimetre scale). a, Adhesion. The top left schematic shows the set-up for adhesion measurement. The test surface is brought in contact with FLIPS at a speed of 0.1 mm s^{-1}, held still for 10 s and then lifted at 1 mm s^{-1}. An example of the resulting force–distance curve is shown at top right. The test surface in this example is a smooth PTFE. The photographs at the bottom demonstrate switchable adhesion. **b**, Pumping. The top is a schematic of the pumping mechanism. The pair of magnets moves the ferrofluid plug along the FLIPS pipe, which in turn drives the liquid (blue) flow inside the pipe. The bottom photographs show that an ethanol solution of rhodamine B is pumped from the right vial to the left vial. **c**, A green algae biofilm is removed by swirling a magnet under FLIPS. Note that the ferrofluids used in these experiments are not toxic, as indicated by the green colour of the algal biofilm.

follow the circular motion of the pairs of magnets, resulting in continuous pumping enabled by a simple rotation, without the need for any complex sequential movement of magnets, as described previously[27]. Last, using a non-toxic fluorocarbon-based ferrofluid, we demonstrate biofilm removal from FLIPS (Fig. 4c). We cultured the algae biofilm on FLIPS under quiescent conditions for about a week, and then moved a magnet under the FLIPS and used the macroscopic protuberance to disrupt the green algae biofilm and detach it from the surface (Supplementary Video 10).

The multiscale topographical response of FLIPS not only has intriguing spatial and temporal fluid dynamics features, but also provides a wide range of interesting phenomena and novel functions when interfaced with other solids and liquids. Our results suggest that FLIPS allows much more diverse combinations of functional capabilities than surfaces that have only a simple, single-scale topographical response. The applications that we have demonstrated—new forms of hierarchical colloidal self-assembly, manipulation and transport of non-magnetic matter in a magnetic field enabled by topography-induced hydrodynamic and capillary forces, controlled formation of droplet clusters of well-defined size, and switchable and directional adhesion and friction—are a small representative subset of these capabilities. Nonetheless, this subset goes well beyond recently reported examples of manipulation of magnetic droplets, controlled wettability and slippery ice-phobic surfaces[5,12–14,18,21,27]. We emphasize the customizability of topographical reconfigurations of FLIPS, which can be tuned by changing the magnetic field, the ferrofluid and, especially, the geometry and orientation of the confining microstructured surface, which itself can be made dynamic by using flexible microstructures[9,28]. The mechanistic insights gained in understanding the physical forces that govern these phenomena and the ensuing functions can be readily applied to other technologically relevant developments: for example, to explore the influence of surface topography on turbulent flow[29], to explore the use of the magneto-caloric effect[15] to manage heat transfer with its surrounding system[30], and to explore the minimization of the pump for novel microfluidics platforms[31]. In addition, we expect the concept of dynamically reconfigurable multiscale topographies to find uses in biology, such as for controlling and stimulating living matter simultaneously at multiple length scales[32–34]. We anticipate that FLIPS and their future developments will benefit areas such as responsive coatings, digital microfluidics and the interfacing of biological tissues with dynamic materials.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0250-8.

Received: 17 July 2017; Accepted: 9 May 2018;
Published online: 25 June 2018

1. Vaia, R. & Baur, J. Adaptive composites. *Science* **319**, 420–421 (2008).
2. Kuroki, H., Tokarev, I. & Minko, S. Responsive surfaces for life science applications. *Annu. Rev. Mater. Res.* **42**, 343–372 (2012).
3. Xia, F. & Jiang, L. Bio-inspired, smart, multiscale interfacial materials. *Adv. Mater.* **20**, 2842–2858 (2008).
4. Yoshida, R. Self-oscillating gels driven by the Belousov–Zhabotinsky reaction as novel smart materials. *Adv. Mater.* **22**, 3463–3483 (2010).
5. Reddy, S., Arzt, E. & del Campo, A. Bioinspired surfaces with switchable adhesion. *Adv. Mater.* **19**, 3833–3837 (2007).
6. van Oosten, C. L., Bastiaansen, C. W. M. & Broer, D. J. Printed artificial cilia from liquid-crystal network actuators modularly driven by light. *Nat. Mater.* **8**, 677–682 (2009).
7. Liu, D., Liu, L., Onck, P. R. & Broer, D. J. Reverse switching of surface roughness in a self-organized polydomain liquid crystal coating. *Proc. Natl Acad. Sci. USA* **112**, 3880–3885 (2015).
8. Liu, D. & Broer, D. J. New insights into photoactivated volume generation boost surface morphing in liquid crystal coatings. *Nat. Commun.* **6**, 8334 (2015).
9. Sidorenko, A., Krupenkin, T., Taylor, A., Fratzl, P. & Aizenberg, J. Reversible switching of hydrogel-actuated nanostructures into complex micropatterns. *Science* **315**, 487–490 (2007).
10. He, X. M. et al. Synthetic homeostatic materials with chemo-mechano-chemical self-regulation. *Nature* **487**, 214–218 (2012).
11. Aizenberg, J., Hatton, B., Yao, X., Aizenberg, M. & Wang, W. Dynamic and switchable slippery surfaces. US patent 9,683,197 (2017).
12. Khalil, K. S., Mahmoudi, S. R., Abu-dheir, N. & Varanasi, K. K. Active surfaces: Ferrofluid-impregnated surfaces for active manipulation of droplets. *Appl. Phys. Lett.* **105**, 0416604 (2014).
13. Tian, D. et al. Fast responsive and controllable liquid transport on a magnetic fluid/nanoarray composite interface. *ACS Nano* **10**, 6220–6226 (2016).
14. Irajizad, P., Hasnain, M., Farokhnia, N., Sajadi, S. M. & Ghasemi, H. Magnetic slippery extreme icephobic surfaces. *Nat. Commun.* **7**, 13395 (2016).
15. Rosensweig, R. E. *Ferrohydrodynamics* (Dover Publications, Mineola, 1997).
16. Rosensweig, R. E. Magnetic Fluids. *Sci. Am.* **247**, 136–145 (1982).
17. Odenbach, S. (ed) *Lecture Notes in Physics: Colloidal Magnetic Fluids: Basics, Development and Application of Ferrofluids* (Springer, Berlin, 2009).
18. Timonen, J. V. I., Latikka, M., Leibler, L., Ras, R. H. A. & Ikkala, O. Switchable static and dynamic self-assembly of magnetic droplets on superhydrophobic surfaces. *Science* **341**, 253–257 (2013).
19. Torres-Díaz, I. & Rinaldi, C. Recent progress in ferrofluids research: novel applications of magnetically controllable and tunable fluids. *Soft Matter* **10**, 8584–8602 (2014).
20. Timonen, J. V. I., Demirörs, A. F. & Grzybowski, B. A. Magnetofluidic tweezing of nonmagnetic colloids. *Adv. Mater.* **28**, 3453–3459 (2016).
21. Yao, X. et al. Adaptive fluid-infused porous films with tunable transparency and wettability. *Nat. Mater.* **12**, 529–534 (2013).
22. Schellenberger, F. et al. Direct observation of drops on slippery lubricant-infused surfaces. *Soft Matter* **11**, 7617–7626 (2015).
23. Parsegian, V. A. *Van der Waals Forces: a Handbook for Biologists, Chemists, Engineers, and Physicists* (Cambridge Univ. Press, Cambridge, 2006).
24. Israelachvili, J. N. *Intermolecular and Surface Forces* 3rd edn, Ch. 13 (Academic Press, Burlington, 2011).
25. Reiter, G. et al. Thin film instability induced by long-range forces. *Langmuir* **15**, 2551–2558 (1999).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

26. Daniel, D., Timonen, J. V. I., Li, R., Velling, S. J. & Aizenberg, J. Oleoplaning droplets on lubricated surfaces. *Nat. Phys.* **13**, 1020–1025 (2017).
27. Hatch, A., Kamholz, A. E., Holman, G., Yager, P. & Bohringer, K. F. A ferrofluidic magnetic micropump. *J. Microelectromech. Syst.* **10**, 215–221 (2001).
28. Grinthal, A. & Aizenberg, J. Adaptive all the way down: building responsive materials from hierarchies of chemomechanical feedback. *Chem. Soc. Rev.* **42**, 7072–7085 (2013).
29. McKeon, B. J., Sharma, A. S. & Jacobi, I. Experimental manipulation of wall turbulence: A systems approach. *Phys. Fluids* **25**, 031301 (2013).
30. Park, K.-C. et al. Condensation on slippery asymmetric bumps. *Nature* **531**, 78–82 (2016).
31. Pamme, N. Magnetism and microfluidics. *Lab Chip* **6**, 24–38 (2006).
32. Verho, T. et al. Reversible switching between superhydrophobic states on a hierarchically structured surface. *Proc. Natl Acad. Sci. USA* **109**, 10210–10213 (2012).
33. Epstein, A. K., Wong, T. S., Belisle, R. A., Boggs, E. M. & Aizenberg, J. Liquid-infused structured surfaces with exceptional anti-biofouling performance. *Proc. Natl Acad. Sci. USA* **109**, 13182–13187 (2012).
34. Sutton, A. et al. Photothermally triggered actuation of hybrid materials as a new platform for *in vitro* cell manipulation. *Nat. Commun.* **8**, 14700 (2017).

Acknowledgements This work is supported by DOE under award number DE-SC0005247 (experiment) and by the NSF under award number DMREF-1533985 (theory). This work was performed in part at the Center for Nanoscale Systems at Harvard University (CNS), which is supported by the NSF under award number ECS-0335765. We thank J. C. Weaver for assistance with 3D printing, M. Khan for assistance with preparing micro-textured silicon master, P. Kim and J. Alvarenga for assistance with force measurements, K. Davey at American Electromechanics for assistance with designing the electromagnet, and N. Vogel, I. Morrison, Y. Hu, P. Kim, D. Daniel, M. Kreder and O. Ahanotu for discussions. W.W. thanks the Humboldt foundation for a fellowship. J.V.I.T. was supported by the European Commission through the Seventh Framework Programme (FP7) project DynaSLIPS (project number 626954). D.-M.D., M.S. and W.W. are supported by the Max Planck Society.

Author contributions J.A., W.W. and J.V.I.T. designed the experiments. W.W. performed the experiments on the dynamics of micro-topography, droplet manipulation, pumping and wrapping-layer-related calculations. J.V.I.T. performed experiments with magnet arrays and manipulation of colloids. C.T.Z. fabricated the microstructures. D.-M.D. measured the ferrofluid overlayer thickness. D.-M.D., M.S. and W.W. performed the adhesion and friction measurements. S.Ko. and J.V.I.T. performed the biofilm experiments. L.M., A.C., W.W. and J.V.I.T. performed the scaling analysis. A.G. assisted in the design of experiments. S.Ke. and J.C. assisted in droplet experiments. R.T.B. assisted in the pump design. T.-S.W., B.H. and S.H.K. contributed to the initial conception of the project. J.A., W.W., J.V.I.T., A.G. and A.C. wrote the manuscript. J.A. supervised the research.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0250-8.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0250-8.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to J.A.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Materials. Hexamethyldisiloxane (HMDS) was obtained from Microchem. Positive i-line photoresist megaposit SPR700-1.0 and developer microposit MF CD-26 were obtained from Microchem. Trichloro(1H,1H,2H,2H-perfluorooctyl) silane (13F-silane) was obtained from Sigma-Alrich. Sylgard 184 silicone (PDMS) elastomer kit was obtained from Dow Corning. Ultraviolet (UV)-curable epoxies (EPO-TEK OG142 and OG178) were obtained from Epoxy Technology. Polytetrafluoroethylene (PTFE) membranes with average pore sizes of 1 μm, 10 μm or 20 μm were obtained from Sterlitech and used as received. Aeos extruded microporous ePTFE tubing (ID 0.0775″ and wall thickness 0.034″, internodal distance 5–60 μm) was obtained from Zeus and used as received. Neodymium magnets (surface fields 1,000–5,000 G) were obtained from K&J magnets or Supermagnete. Alnico magnets were obtained from All Magnetics. Fluorocarbon-based ferrofluids were obtained from Ferrotec. The perfluoropolyether oil used in the dilution of ferrofluids was Dupont Krytox 100, obtained from 3 M. Extended Data Table 1a summarizes the physical properties of the two types of fluorocarbon-based ferrofluid used in our experiments as well as the properties of Krytox 100. Reagent plus grade dodecane was obtained from Sigma-Aldrich.

The silicone-oil-based ferrofluid was synthesized by using the coprecipitation method. In brief, 720 ml of milli-Q water was mixed with 21.6 g of iron(III) chloride hexahydrate and 11.2 g of iron(II) sulfate heptahydrate, followed by coprecipitation by adding 80 ml of ammonium hydroxide (28%–30%). The iron oxide nanoparticles formed were functionalized by adding 25 g of monocarboxydecyl-terminated polydimethylsiloxane (Gelest MCR-B12) and allowing it to react overnight. Particles were purified by sedimenting them with a strong magnet, followed by removal of the aqueous supernatant and adding 400 ml of acetone to redisperse the particles. Particles were further magnetically sedimented, acetone supernatant discarded, and 250 ml of toluene added. Particles were again magnetically sedimented and toluene supernatant removed. Finally, 50 ml of toluene was added and any remaining traces of acetone were removed by heating to 60 °C for a few hours. The resulting volume of the ferrofluid in toluene was approximately 80 ml. Finally, iron oxide particles were transferred to silicone oil by mixing the nanoparticle dispersion in toluene with silicone oil and evaporating the toluene. Fluorocarbon-based and silicone-oil-based ferrofluids provide qualitatively similar results.

The preparations of different FLIPS and additional materials for specific experiments are described in detail below.

Microstructured substrates. Microstructured silicon substrates (masters) were fabricated using the Bosch process and replicated using soft lithography based on previous protocols[35,36]. In brief, adhesion promoter (HMDS) and positive i-line photoresist (SPR700-1.0, about 2 μm height) were spin-coated onto a clean wafer, soft-baked at 95 °C for 60 s, patterned using a direct write laser tool (Heidelberg, Maskless Aligner, 405 nm laser at about 125 mJ cm^{-2} s^{-1}), hardened at 115 °C for 60 s, and developed in the developer CD-26 for around 90 s. The photoresist-patterned silicon wafer was then ion-etched under optimized Bosch conditions (SPTS Technologies) to nominal height and rinsed in acetone/isopropanol to remove residual photoresist. The resulting silicon microstructures were treated with plasma and passivated with 13F-silane under vacuum for more than 6 h. The silicon masters were then used to prepare polydimethylsiloxane (PDMS) negative moulds. PDMS prepolymer (base-to-hardener ratio of 10:1, wt/wt) was poured onto the master, cured at 70 °C for 2 h, and peeled from the masters to obtain negative moulds. PDMS negative moulds were stored in 13F-silane vapour environment for more than 3 h. The PDMS moulds were then used to create an epoxy positive replica. A few millilitres of epoxy (EPO-TEK OG142 or OG178) were cast at the centre of a Petri dish using a plastic pipette. A PDMS mould was then carefully placed over the epoxy to prevent bubbles from forming between PDMS and the epoxy layer. To obtain epoxy replicas with areas smaller than that of the PDMS moulds, a few drops of epoxy (less than 1 ml) were cast on PDMS mould, and a clean glass slide (rinsed with acetone, isopropanol and water, and blow-dried with nitrogen) was placed slowly over the epoxy, followed by curing under UV light (Bio-Link 365, Vilber) for 20 min. The PDMS mould was then peeled off slowly to give the microstructured positive epoxy replica. Any variations of the procedures for preparing positive epoxy replicas are described in the relevant sections.

Dynamics of the micro-topographical response of FLIPS. The silicon masters are prepared on 10-cm wafers (area of about 80 cm^2). Epoxy replicas were made inside a 13.5-cm-diameter Petri dish. To prepare FLIPS, 0.4 ml of fluorocarbon-based ferrofluid 1 was dropped onto the epoxy microstructures and spread over the substrate surface with a magnet. To vary the overlayer thickness, the FLIPS sample was spun at 500 r.p.m. for a period of 1–5 min, and the overlayer thickness was measured directly using a force probe (see below). At the beginning of each experiment, magnets were fixed to the back of the Petri dish using double-sided tape. For the data presented in Fig. 1d and Extended Data Fig. 2c, a 10-mm-diameter and 40-mm-tall N45 neodymium iron boron (NdFeB) magnet was used. In other control experiments, 0.5-inch-diameter and 0.5-inch-tall N52 NdFeB or 4-mm-diameter and 7-mm-tall N45 NdFeB was used. The dynamics of the micro-topographical response was recorded using an angled illumination technique (see below). The videos were then analysed using a custom Matlab code. The colour thresholding was performed on V values in the HSV colour space. In the experiments with the spiral pattern, 4-mm-diameter and 7-mm-tall N45 NdFeB was used.

Angled illumination technique (Extended Data Fig. 1b). A cold LED line light (Zeiss CL 6000) was used to illuminate the FLIPS sample at an incidence angle of approximately 30°. A video camera (Sony HDR CX900) was positioned directly above the FLIPS sample to record the process. Scattered light from the micro-topographical regions greatly increased the brightness of the region and facilitated the video-processing step of the analysis.

Force-probe technique to measure the thickness of the ferrofluid overlayer (Extended Data Fig. 1c). A spherical glass probe was lowered on the FLIPS surface at a controlled speed of 10 μm s^{-1}. Force was recorded at 1-ms intervals using a load cell (Transducer Techniques). The probe was programmed to stop when a contact force of 50 mN was reached. The thickness of the ferrofluid overlayer was then determined from the resulting force–displacement curve as the distance between capillary snap-in and the position where the probe touches the top of the microstructure.

Measurement of 3D profiles of the ferrofluid–air interface. The 3D profiles of the ferrofluid–air interface were measured using a 3D laser-scanning confocal microscope (Keyence) with a 100× long-working-distance objective. For the data on the evolution of 3D profiles over time and the change in profiles along the *x* axis (Fig. 1c, Extended Data Fig. 2), the microstructured substrates used were epoxy replicas of pattern 1 fabricated on a 1 inch × 3 inch glass slide, with the microchannels aligned parallel to the long edge of the glass slide, and fluorocarbon-based ferrofluid 1 was used. A 4-mm-diameter and 7-mm-tall N45 NbFeB magnet was used. For the data on fine-tuning the balance between magnetic and capillary pressures (Extended Data Fig. 3), the microstructured substrates were epoxy replicas of pattern 1 fabricated on a 1 inch × 3 inch glass slide, with the microchannels aligned perpendicular to the long edge of the glass slide, or epoxy replicas of pattern 5. Fluorocarbon-based ferrofluid 1 was used. The magnets used were either 10-mm-diameter and 7-mm-tall N45 NdFeB magnets or 0.5-inch-diameter and 0.5-inch-tall Alnico magnets with a surface field of 0.041 T.

Topographical responses to magnet arrays. FLIPS were prepared with epoxy replicas of pattern 4 on a 1-mm-thick glass slide. Fluorocarbon-based ferrofluid 1 was used to induce both micro-topographical and macro-topographical responses. Multiple small cylindrical magnets were first embedded into an acrylic sheet. The sheet was drilled with a hexagonal pattern of circular holes with a CO_2 laser-cutting system (Versalaser). The diameter of the holes was chosen to be slightly less than the diameter of the cylindrical magnets (K&J Magnetics, NdFeB, diameter 1/16″) so that individual magnets could be mounted by simply pressing them into the holes without using any adhesives. The magnet array was placed on top of a white-LED panel light and the FLIPS was lowered on top of the magnet array (to direct contact with the magnets). The propagation of the depletion pattern was imaged immediately with a digital camera (Panasonic DMC-GH4) equipped with a macro lens (Olympus Zuiko 60 mm).

Applications in colloidal assembly and transport. *Assembly of non-magnetic colloidal particles on FLIPS.* The microplate array (pattern 4) was created on the top of a standard 1 inch × 3 inch glass slide by replication moulding of a silicon master. In brief, PDMS (1:10) was cast on the silicon master, degassed under vacuum, cured at 70 °C and peeled off. The resulting negative mould was filled with UV-curable photopolymer (Norland Optical Adhesive 61) and pressed against the glass slide. Photopolymer was cured under UV light (Dymax 2000-EC) for about 1 min and the PDMS mould was peeled off. A liquid reservoir for the colloidal dispersion was created by attaching an aluminium washer ring (inner diameter of about 20 mm, height of a few millimetres) on top of the microstructures with the same UV-curable photopolymer. The microstructures were then lubricated by adding a droplet of non-diluted fluorocarbon-based ferrofluid (Ferrotec) or silicone-oil-based ferrofluid to completely cover the microplates. After the microplates were covered with ferrofluid, the ferrofluid could be locally depleted from the microstructures by applying a local magnetic field with a small permanent magnet. The reservoir was filled with a colloidal dispersion consisting of non-magnetic carboxylate-modified fluorescent (rhodamine-B-marked) melamine microparticles 10 μm in diameter (Sigma-Aldrich). Colloidal particles were allowed to settle down for a few hours after which they were imaged with an upright confocal microscope (Zeiss LSM710) with water-dipping objectives (40×/1.0 and 10×/0.3). Depending on the state of the FLIPS surface, the colloids were found to assemble randomly on the 2D plane (flat ferrofluid–water interface; Fig. 2a) or to follow the underlying microstructure, forming particle chains (partially depleted structures in micro-topographical area; Fig. 2b).

Hierarchical assembly. More complicated colloidal assemblies were created by using magnetically patterned FLIPS surfaces. The FLIPS substrates were created in the same way as described in Methods subsection 'Assembly of non-magnetic colloidal particles on FLIPS', with the exception that a hexagonal nickel grid

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

(Gilder hexagonal grid 100 mesh, Ted Pella) was placed between the glass slide and the PDMS mould filled with photopolymer before curing it with UV light, resulting in a microplate array with a magnetic grid under it (Fig. 2h). If no microplates were desired (that is, plain flat surface), then a flat PDMS block was used (Fig. 2g). Lubrication and colloidal assembly were carried out as described in the previous section.

Transport of non-magnetic colloids on FLIPS. Magnetically induced motion of the non-magnetic colloids was demonstrated by applying a horizontal magnetic-field gradient on the colloids that had settled on a partially depleted micro-topographical area of FLIPS. A gradient was created by placing an NdFeB magnet (cube, about 1-inch sides) next to the substrate and imaging was done with a long-working-distance zoom microscope (Zeiss AxioZoom). Motion of the colloids was stopped by removing the magnet (Fig. 2c–f).

Droplet flow control. *Effect of micro-topography.* FLIPSs were prepared with microstructured epoxy replicas of pattern 4 for the first part of the demonstration and of pattern 1 for the second part of the demonstration. For the second part of the demonstration, the angled illumination technique was used to visualize the micro-topographical region directly. Fluorocarbon-based ferrofluid 1 was used to induce a micro-topographical response. FLIPS was placed at a tilt angle of 15°–30° to induce droplet flow due to gravity. A stack of permanent magnets (three to four 0.5-inch-diameter and 0.5-inch-tall N52 NdFeB for the first demonstration; five 4-mm-diameter and 7-mm-tall NdFeB for the second demonstration) was placed on the top of a linear actuator (Firgelli L12P Linear Actuator w/ LAC); the use of a stack of magnets helps to extend the magnetic field along the z direction. The linear actuator has an actuation range of 10 cm and was used to adjust the relative vertical position of the magnets with respect to FLIPS precisely. During an experiment, the magnets were first brought close to FLIPS, with a separation distance of about 1 mm, to induce a micro-topographical response. After the micro-topographical region was formed, usually after about 30 min, a droplet of water (5–10 μl) was placed in the area above the micro-topographical region so that it would begin to slide down owing to gravity. Once the droplet moved into the micro-topographical region, it slowed and eventually stopped. The magnets were lowered to a distance 10–20 mm below the FLIPS so that the ferrofluid could flow back to the micro-topographical region, allowing the droplet to begin sliding again.

Effect of macro-topography. FLIPSs were prepared with microstructured epoxy replicas of pattern 4. Fluorocarbon-based ferrofluid 2 was diluted to 4% by volume with Krytox 100 to reduce the magnetic pressure and suppress the micro-topographical response. A stack of permanent magnets (three to four 0.5-inch-diameter and 0.5-inch-tall N52 NdFeB) was placed on the top of a linear actuator (Firgelli L12P Linear Actuator w/ LAC). The linear actuator has an actuation range of 10 cm and was used to adjust the relative vertical position of the magnets with respect to FLIPS precisely. Furthermore, the linear actuator was attached to a translational stage that was adapted and modified from the x-stage of the 3D printer RepRapPro Huxley. The x-stage and the linear actuator were both controlled through a PC interface. In the experiment, the magnets were first brought close to the FLIPS, with a separation distance of about 1 mm, to form the macro-topographical protuberance. A syringe equipped with a needle, pumped by a syringe pump, was positioned at the top of the FLIPS to introduce a flow of droplets. The magnets were then translated in the horizontal direction to move the macroscopic protuberance into the track of the droplet flow to induce droplet pinning. The magnets were then lowered successively to the desired distance from FLIPS (1–5 mm) to induce different assembling behaviours of droplets.

Control of the wrapping layer formation and droplet mixing. *Interfacial tension measurements.* Surface and interfacial tension measurements of various liquids used in the study were performed on KSV Instruments's CAM 101 system using the pendant drop method. Deionized water with a resistivity of more than 18.2 MΩ cm was collected from Millipore's Milli-Q water system. Dodecane of analytical standard grade (AS, >99.8%) and reagent plus grade (RP, >99%) were obtained from Sigma-Aldrich; dodecane of 99+% grade was obtained from Alfa Aesar.

Formation and removal of wrapping layers. FLIPS were prepared on porous PTFE membranes with 1-μm pore size to suppress the micro-topographical response. Fluorocarbon-based ferrofluids and their 4%-by-volume dilution by the fluorocarbon oil Krytox 100 were used in the experiments to study the wrapping layer, to avoid mixing between the ferrofluid and the content of the droplets. For the experiments shown in the Supplementary Video 5, ferrofluid 2 was used. A camera (Cannon Repel T2i) was used to record the experiment. For the formation of the wrapping layer, a water droplet was placed slowly on the FLIPS in air. For the removal of the wrapping layer, the FLIPS and the water droplet were placed in a beaker, into which a hydrocarbon was added slowly.

Mixing droplets with wrapping layers. FLIPS were prepared on porous PTFE membranes with 1-μm pore size to suppress the micro-topographical response. Fluorocarbon-based ferrofluid 2 diluted to 4% by volume with Krytox 100 was used. The electromagnet in this experiment consisted of 286 turns of a heavy film #14 AWG wire wrapped around a 2.61-cm-long and 0.9-cm-diameter iron-cobalt-vanadium Hiperco 50 alloy core (custom fabrication from Dura Magnetics). This electromagnet was designed to operate between 1 Hz and 550 Hz, producing non-hysteretic sinusoidal magnetic fields of more than 0.2 T when powered using an APS-1102 programmable AC/DC power source (Instek America). In the experiment, two droplets of colloidal suspensions (10 μl each) were placed near the centre of the electromagnet, one to the left and the other to the right. The right one was dyed with rhodamine B to induce visual differences. Upon turning on the electromagnets, the droplets move and mix on the top of the alloy core of the electromagnet. An LED was used to provide enough illumination for video recording (Cannon Repel T2i).

Mixing droplets without wrapping layers. FLIPS were prepared on porous PTFE membranes with 1-μm pore size to suppress the micro-topographical response. Fluorocarbon-based ferrofluid 2 diluted to 4% by volume with Krytox 100 was used. The FLIPS was placed in a plastic Petri dish, which was then filled with dodecane. Droplets (about 10 μl) of aqueous solution of sodium bicarbonate and hydrogen chloride were then placed on the FLIPS with a micro-pipette. A permanent magnet was used to induce a macroscopic response to push one droplet towards the other and mix them.

Adhesion and friction experiments. *Adhesion measurements on PTFE-membrane-based FLIPS.* FLIPS were prepared with a porous PTFE membrane with 20-μm pore size and fluorocarbon-based ferrofluids 1 and 2 without dilution to induce both micro-topographical and macro-topographical responses. As a result, the size of the macro-topographical protuberance and its ability to separate two surfaces spontaneously was influenced not only by the ferrofluid overlayer but also by the amount of ferrofluid extracted from the porous membrane. Adhesion measurements were performed on an Instron 5566 electromechanical testing system. A plastic sample holder was designed and 3D-printed to fit the gripper of the instrument and to provide a flat top surface with an area of 1 inch^2. The porous PTFE membrane was first attached to the top of the 3D-printed sample holder with double-sided tape and then 150 μl of ferrofluid was added. A glass cover slide, a polytetrafluoroethylene Teflon sheet (McMaster Carr, 0.015-inch thick, one side adhesive-ready) and a slippery liquid-infused porous surface (SLIPS) were attached to another sample holder. The FLIPS was positioned at the bottom and the test surface at the top. During testing, the FLIPS stayed stationary, and the test surface was first brought down at a speed of 0.1 mm s^{-1} to be in contact with the FLIPS, held still for 10 s and then retracted at 0.01 mm s^{-1}, 0.1 mm s^{-1} or 1 mm s^{-1}. Position $l = 0$ corresponds to the position where the top surface was pressed slightly against the FLIPS without squeezing ferrofluid out of the FLIPS. For each data point, a minimum of five measurements were performed.

Adhesion measurements on channel-like microstructure-based FLIPS. The microstructure of pattern 1 was used to demonstrate the anisotropic properties of FLIPS. The substrates were prepared with glass slide backing. Epoxy (EPO-TEK OG178) was cast onto the PDMS mould and then the glass slide was placed onto the epoxy, creating a homogeneous thin epoxy layer between the PDMS mould and the glass slide, and cured under UV light. The microstructured epoxy resin was then peeled off the mould, infiltrated with 100 μl of fluorocarbon-based ferrofluid 1 and spun at 500 r.p.m. for 60 s before being attached to another sample holder. The overlayer thickness was estimated to be about 20 μm. The adhesion measurements were performed on a customized set-up. The adhesion set-up was built on an inverted optical microscope (Axio Observer A1, Zeiss) with a video camera (Grasshopper3, Point Grey Research), enabling the recording of the contact interface. The adhesion force was measured by a sensitive load cell (GSO-25 and −1 K, Transducer Techniques) mounted on a computer-controlled high-precision piezo motion stage (LPS-65 2″, Physik Instrumente) in the z direction, with a resolution of 5 nm and a maximum velocity of 10 mm s^{-1}. Fine positioning in the x and y directions was done by a manual xy stage (NFP-2462CC, Positionierungstechnik Dr Meierling) and tilt correction was adjusted by two goniometers (M-GON65-U, Newport). Motion control of the piezo stages and data acquisition were performed using a customized Linux code (Ubuntu, Canonical). The program enabled control over preload, velocity, displacement in the x and z directions, and contact time. The load cell was linked to the computer via a signal conditioner (BNC-2110, National Instruments) and the voltage signal from the force measurement was transferred through a data acquisition board (PCIe-6259, National Instruments). A brass holder with attached plastic disk was prepared to attach the PDMS probe to the load cell. The circular plastic disk with 4-mm diameter and 1-mm thickness was laser-cut from a plastic plate and attached to the brass holder by a silicone adhesion promoter (Sil-Poxy, Smooth-On). The FLIPS was positioned at the bottom and the holder on top. Sylgard 184 prepolymer and curing agent with weight ratio of 10:1 were mixed, degassed and cast on a glass plate, and a thin film with 500-μm thickness was created by a film applicator (Multicator 411, Erichsen). The sample was cured in a vacuum oven at 90°C for 1 h. Square PDMS probes with 5-mm side length were cut from a 500-μm-thick PDMS film and placed onto the FLIPS surface to ensure alignment. A vinylsiloxane polymer (Flexitime Medium Flow, Heraeus Kulzer) was used to bond the PDMS probe to the holder. The holder was

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

inked into the uncured vinylsiloxane polymer film, moved and contacted with the backside of the aligned PDMS probe with 50-mN compressive load. After 3 min, the vinylsiloxane polymer was polymerized and the probe bonded to the holder. During the adhesion testing, the FLIPS stayed stationary; the probe approached the surface at 50 μm s^{-1} and was first brought in contact with a preload of 50 mN. After a contact time of 10 s, the probe was retracted at a speed of 10 μm s^{-1}, 100 μm s^{-1} or 1,000 μm s^{-1} until the probe was detached from the FLIPS. The probe was cleaned after each measurement with a particle-free tissue and isopropanol to remove accumulated ferrofluid. Pristine and representative positions of the different regions (H_0 off, R1, R2 and R3) were selected for each measurement. The ferrofluid was respread over the substrate surface with a magnet and the initial conditions were restored, and all pristine and representative positions were measured. The experiments were conducted in a temperature- and humidity-controlled laboratory with the conditions kept at 20–25 °C and 25%–35%, respectively. For each data point, a minimum of five measurements were performed.

Friction measurements on channel-like microstructure-based FLIPS. The FLIPS samples were prepared in the same way as in Methods subsection 'Adhesion measurements on channel-like microstructure-based FLIPS'. Friction measurements were performed on the same customized set-up. The load cell was attached to the *z*-direction piezo motion stage perpendicular to the FLIPS. A plastic holder was designed and laser cut to attach the PDMS film probe to the load cell and to provide aligned configuration. Sylgard 184 prepolymer and curing agent with a weight ratio of 10:1 were mixed, degassed and cast on a glass plate, and a thin film with 500-μm thickness was created by a film applicator (Multicator 411, Erichsen). The sample was cured in a vacuum oven at 90 °C for 1 h. The PDMS film probe with length of 25 mm was cut from a 500-μm-thick PDMS film. A 1-cm^2 PDMS piece was attached to the lower side of the film probe by a thin layer of silicone adhesion promoter to provide a constant contact area. The PDMS film probe was bonded to the plastic holder by a silicone adhesion promoter. The positioning in the *x* and *y* directions was done by a manual *xy*-stage and tilt was corrected by two goniometers to ensure parallel shear. During testing, the film probe stayed stationary and the FLIPS was sheared parallel at a constant velocity. The film probe approached the surface at 1,000 μm s^{-1} and was brought in contact with the substrate. A constant load of 10 mN or 100 mN was applied. After a contact time of 10 s, the FLIPS was sheared over a distance of 5 mm and 10 mm with a velocity of 100 μm s^{-1} and 1,000 μm s^{-1}. The probe was cleaned after each measurement with a particle-free tissue and isopropanol to remove accumulated ferrofluid. Pristine and representative positions of the different regions (H_0 off, R1, R2 and R3) were selected for each measurement. The ferrofluid was respread over the substrate surface with a magnet and the initial conditions were restored, and all pristine and representative positions were measured. The experiments were conducted in a temperature- and humidity-controlled laboratory, with the conditions kept at 20–25 °C and 25%–35%, respectively. For each data point, a minimum of five measurements were performed.

Demonstration of switchable adhesion. FLIPS were prepared on porous PTFE membranes with 20-μm pore size. Ferrofluid 2 was used to induce macro-topographical and micro-topographical responses. As a result, the size of the macro-topographical protuberance and its ability to separate two contacting surfaces spontaneously was influenced not only by the ferrofluid overlayer but also by the amount of ferrofluid extracted from the porous membrane. 3D models were created in Rhinoceros 3D with the aid of the Grasshopper plugin and were printed on Objet Connex 500. Before the experiment, the PTFE membrane was attached to the top of the white holding frame and the ferrofluid was added and spread using a magnet. A beaker was then placed insider the holding frame and the top handle was placed on the FLIPS. Three separate 0.5-inch-diameter and 0.5-inch-tall N52 NbFeB magnets were introduced manually to generate a topographical response. The macro-topographical protuberance (the size of which was determined not only by the ferrofluid overlayer but also by the amount of ferrofluid extracted from the microporous membrane, that is, the micro-topographical response) creates a gap that reduces the adhesion between the top handle and the bottom frame. The magnets were later removed with a magnetic stainless-steel tweezer.

Demonstration of switchable friction. FLIPS were prepared on porous PTFE membranes with 20-μm pore size. Ferrofluid 2 was used to induce macro-topographical and micro-topographical responses. As a result, the size of the macro-topographical protuberance and its ability to separate two contacting surfaces spontaneously was influenced not only by the ferrofluid overlayer but also by the amount of ferrofluid extracted from the porous membrane. 3D models were created in Rhinoceros 3D with the aid of the Grasshopper plugin and were printed on Objet Connex 500. A stepper motor (Lin Engineering, 3518M-07) was used to provide the rotation. Arduino Uno microcontroller with Adafruit motor shield was used to control the rotation of the stepper motor and interface it with a laptop PC. The 3D-printed set consisted of a bottom spinning wheel that was attached to the shaft of the stepper motor, a top spinning wheel and a top stationary wheel. The PTFE membrane was first attached to the top surface of the bottom spinning wheel and the ferrofluid was added to the middle of the membrane. The top spinning wheel was then pressed against the bottom spinning wheel to spread the ferrofluid uniformly throughout the PTFE membrane. Two stacks of plastic Petri dishes and several plastic sheets were used to adjust the height of the top stationary wheel so it could constrain the position of the top spinning wheel. Four small black parts were attached to the top spinning wheel as a visual indicator of spinning and to adjust the overall friction during spinning. At the centre of the top spinning wheel, a 0.5-inch-wide hole was created to accommodate a stack of two 0.5-inch-diameter and 0.5-inch-tall N52 NbFeB magnets. During the demonstration, the movement of the stepper motor rotates the bottom spinning wheel, which rotates the top spinning wheel without magnets. When the magnets were placed in the hole in the top spinning wheel, the macro-topographical protuberance (the size of which was determined by not only the ferrofluid overlayer but also by the amount of ferrofluid extracted from the microporous membrane, that is, the micro-topographical response) creates a gap that reduces the friction between the top and the bottom spinning wheels and stops the rotation of the top spinning wheel. The explanatory animation at the end of Supplementary Video 8 was made in Maya, using the models created in Rhinoceros 3D.

Demonstration of pumping. Aeos extruded microporous ePTFE tubing (effective pore size of about 5–60 μm) and ferrofluid 2 were used to allow the ferrofluid to move in and out of the tubing pore wall. 3D-printed parts were used to construct the rest of the pump. 3D models were created in Rhinoceros 3D with the aid of the Grasshopper plugin and were printed on Objet Connex 500. The stepper motor (Lin Engineering, 3518M-07) was used to provide rotation. An Arduino Uno microcontroller with Adafruit motor shield was used to control the rotation of the stepper motor and interface it with a laptop PC. The 3D-printed set consisted of a flat top plate, a bottom plate with a circular track, two linear tracks to accommodate the movement of the ferrofluid protuberance and the tubing, and two spinning wheels, each with ten slots for 0.5-inch-diameter and 0.5-inch-tall magnets. Before the experiment, the ePTFE tube was first fitted into the track in the bottom plate and the ferrofluid was added at various locations along the circular track to wet the tube. Krytox 100 oil was added to various locations along the linear tracks to provide sealing for the liquid being pumped along the inlet and outlet parts of the tube. The top plate was then used to sandwich the tube and the sandwich structure was then placed on the top of the bottom spinning wheel, inside of which five magnets were positioned in the holes along the perimeter with even spacing. The top spinning wheel was then installed, into which another five magnets were placed. The explanatory animation at the end of Supplementary Video 9 was made in Maya, using models created in in Rhinoceros 3D.

Biofilm studies. *FLIPS preparation.* PTFE membranes with a pore size of 10 μm were adhered to a 50 mm × 75 mm glass slide using a thin layer of semi-cured PDMS (Sylgard 184 silicone elastomer Dow Corning Corporation), enabling the thorough attachment of the membrane to the glass slide without infiltrating the PTFE network with PDMS. After completing the PDMS curing in an oven (70 °C for 4 h), 500 μl of ferrofluid was added to the surface of each PTFE membrane and spread carefully using a ring magnet until all of the membrane area was fully infused with ferrofluid. The ferrofluids used in our study consisted of non-toxic components: iron oxide particles dispersed in fluorocarbon solvents. The non-toxicity of many fluorocarbon solvents is well known and manifested by the fact that they have even been used as the main ingredients in artificial blood substitutes (Fluosol) that are FDA-approved and used with human subjects. Iron oxide nanoparticles are also non-toxic, with some specific types FDA-approved as MRI contrast agents. This is in contrast to ferrofluids based on many other types of carrier fluids and nanoparticles, including nickel and cobalt, which are known to be toxic to biological matter.

Algae cultivation and experimental set-up. The green alga *Chlamydomonas reinhardtii* (UTEX number 89) from the University of Texas Culture Collection was used as a model organism to explore the biofilm disruption or detachment potential of FLIPS systems. *C. reinhardtii* was grown in a Soil Extract (Bristol-medium-based) solution under non-axenic conditions until the stock culture reached a density of approximately 10^7 cells ml^{-1}. This stock culture was diluted with fresh Soil Extract to a 1:5 ratio of stock culture to fresh medium. 80 ml of the diluted culture was added to square (10 cm × 10 cm) Petri dishes containing the prepared FLIPS treatments, allowing the algae to settle on test surfaces. The Petri dishes were then placed under a Sun Blaze T5HO fluorescent light fixture (Sunlight Supply) and were grown under a 16 h–8 h light–dark cycle at 24 °C for seven days until a cohesive green algae biofilm had formed. No negative effects of the ferrofluid on algae growth and biofilm formation were observed when compared with PTFE-only control treatments, and a healthy green algae biofilm remained stably attached to the surface until the magnetic actuation was performed, thus confirming the non-toxic nature of FLIPS.

Biofilm removal. To test the potential of FLIPS surfaces to disrupt or detach the adhered biofilm, a ring magnet was introduced immediately beneath the square Petri dishes containing the fouled FLIPS and the algae medium. The applied

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

magnetic fields then led to a concentration of the ferrofluid beneath the magnet, breaking up the adhered green algae biofilm in the process. Moving the magnet up and down the slide caused a 'ferrofluid wave' to travel over the FLIPS, detaching and concentrating the biofilm. The aggregations of biofilms moved along with the ferrofluid wave but did not mix with the underlying ferrofluid. After removing the magnet, the Petri dishes were shaken gently to see if the biofilm remnants were still associated with the ferrofluid. The biofilm threads were not associated, but instead lifted off the FLIPS and started floating freely into the algae media. By contrast, simple shaking of the Petri dishes without magnetic actuation did not result in any biofilm detachment from the FLIPS surface. Control PTFE treatments that did not contain any ferrofluid showed no response to the magnetic actuations and the green algae biofilms remained firmly attached to the surface of the controls.

Simulations and calculations. Simulations of magnetic fields and magnetic pressures were performed with COMSOL. The data for the ferrofluid *M–H* curve were obtained from Ferrotec. Calculation of the force density was performed with OriginLab. Calculations of van der Waals energy and disjoining pressure were performed with Mathematica.

Code availability. The codes used in analysing videos and other data are available from the corresponding author.

Additional notes on the scaling relations of the micro-topographical response of FLIPS. Using the geometry in Fig. 1b, the pressure gradient in the x direction drives the flow and is resisted by viscous dissipation in the z direction; therefore, using the lubrication approximation[37,38], we have

$$\frac{\partial p_\gamma}{\partial x} = \eta \frac{\partial^2 U}{\partial z^2}$$

where $p_\gamma \approx 2\gamma/d_y$ is the capillary pressure and $U \approx \mathrm{d}L_x/\mathrm{d}t \approx L_x/t$. The left-hand side represents the capillary pressure gradient (which drives the porous-capillary flow) and the right-hand side represents the viscous dissipation. Using the approximations $\partial p_\gamma/\partial x \approx p_\gamma/L_x \approx 2\gamma/d_y L_x$ and $\partial^2 U/\partial z^2 \approx U/h_0^2 \approx L_x/(th_0^2)$ and rearranging the terms gives the scaling relation

$$L_x \approx \left(\frac{2\gamma h_0^2}{\eta d_y}\right)^{1/2} t^{1/2}$$

Putting numerical values in the above equation and setting the units of L_x as millimetres and t as seconds gives $L_x \approx 0.35t^{1/2}$. This derivation gives the same power law as in equation (2). The only difference is the pre-factor.

A second alternative is to use magnetic pressure to replace the capillary pressure term:

$$\frac{\partial |p_\mathrm{m}|}{\partial x} = \eta \frac{\partial^2 U}{\partial z^2}$$

where $|p_\mathrm{m}| \approx \mu_0 M_\mathrm{s} H_0$. This replacement represents the incorporation of the actual driving force—the magnetic pressure gradient—in the formulation of the scaling relation. Arranging terms gives the scaling relation

$$L_x \approx \left(\frac{|p_\mathrm{m}| h_0^2}{\eta}\right)^{1/2} t^{1/2}$$

Using $p_\mathrm{m} \approx 10^4$ Pa and again setting the units of L_x as millimetres and t as seconds gives $L_x \approx 5.6t^{1/2}$.

The derivations in the main text and here all give the same power law, and differ only in the pre-factors. A more accurate model will need to combine these models: near the magnet, it is primarily the magnetic force that drives the flow; far from the magnet, it is the capillary pressure gradient that drives the flow. In addition, these scaling derivations are essentially 2D models; they omit the 3D nature of the geometry of the microstructure. We are currently developing numerical models to take into account all of the considerations stated above.

Additional notes on the formation of the wrapping layer. The estimate of the formation of the wrapping layers based on the calculation of the (initial) spreading coefficients does not correlate with our experiments (Extended Data Fig. 4, Extended Data Table 2). This method is derived from work in the early 1920s[39,40]. It was also the method of choice in recent studies on wrapping layers[22,41]. We adopt the term 'initial spreading coefficient' S_i to denote the value that we calculate[42] and reserve the term 'spreading coefficient' to mean the 'true equilibrium spreading coefficient'[42]. The main source of discrepancy for the S_i approach is that interfacial tensions are very sensitive to trace amounts of surface-active agents, especially for a water–hydrocarbon interface. Any additives such as dyes or even air-borne dust particles can alter the water–hydrocarbon interfacial tension sufficiently to change the signs of S_i.

Our estimate of the formation of the wrapping layer is based on the calculations of long-range van der Waals interactions[23–26,38,43]. The origin of these interactions is different from that of the surface tension (which is due to short-range intermolecular forces) and represents the interaction between the two media across the wrapping layer (hence long-range). If this interaction between the two media is attractive, then the wrapping layer is absent; if this interaction is repulsive, then the wrapping layer is formed. Knowing the sign of this interaction enables us to estimate the stability of the wrapping layer (in other words, the absence or presence of the wrapping layer). From the perspective of thin films, a ferrofluid thin film is stable between water and air because the repulsive van der Waals interaction between air and water tends to thicken it; on the other hand, a ferrofluid thin film is unstable between water and hydrocarbon because the attractive van der Waals interaction between water and hydrocarbon tends to thin it.

We provide below three methods to estimate the sign of these long-range van der Waals interactions, in order of increasing complexity: (1) using combining relations (Extended Data Table 3); (2) using non-retarded Hamaker constants based on Lifshitz theory (Extended Data Table 4); and (3) using Hamaker constants with a relativistic retardation correction based on Lifshitz theory (Extended Data Fig. 5).

Method 1. We estimate the signs of the Hamaker constants A_{132} and van der Waals interaction energy G_{132} using combining relations[24,25] (results are presented in Extended Data Table 3):

$$\begin{aligned} A_{132} &= \left(\sqrt{A_{11}} - \sqrt{A_{33}}\right)\left(\sqrt{A_{22}} - \sqrt{A_{33}}\right) \\ &\approx 24\pi\delta_0^2 \left(\sqrt{\gamma_1^\mathrm{D}} - \sqrt{\gamma_3^\mathrm{D}}\right)\left(\sqrt{\gamma_2^\mathrm{D}} - \sqrt{\gamma_3^\mathrm{D}}\right) \\ G_{132} &= -\frac{A_{132}}{12\pi d^2} \approx -\left(\sqrt{\gamma_1^\mathrm{D}} - \sqrt{\gamma_3^\mathrm{D}}\right)\left(\sqrt{\gamma_2^\mathrm{D}} - \sqrt{\gamma_3^\mathrm{D}}\right) \end{aligned}$$

where A_{ijk} is the Hamaker constant for the interaction between medium i and medium k across medium j, G_{ijk} is the van der Waals interaction energy for the interaction between medium i and medium k across medium j, A_{ii} is the Hamaker constant for the interaction between medium i across vacuum, δ_0 is the atomic cut-off distance (a constant), γ_i^D is the dispersive component of the surface tension of medium i and d is the thickness of the intermediate medium. For hydrocarbons and ferrofluids, γ_i^D is equal to the surface tensions in air; for water, $\gamma_i^\mathrm{D} \approx 22\,\mathrm{mN\,m^{-1}}$ (refs [25,43]).

This calculation used the physical properties of fluorocarbon-based ferrofluids and perfluoropolyether oil (Krytox 100) provided by the manufacturers and the measured surface and interfacial tensions of liquids used in the study (Extended Data Table 1).

Method 2. We estimate the non-retarded Hamaker constants A_{132} and van der Waals interaction energy G_{132} using Lifshitz theory (results are presented in Extended Data Table 4). The calculations are based on the following equation, adapted from ref. [24]:

$$\begin{aligned} A_{\mathrm{total}} &= A_{\nu=0} + A_{\nu>0} \\ &\approx \frac{3}{4}kT\left(\frac{\varepsilon_1-\varepsilon_3}{\varepsilon_1+\varepsilon_3}\right)\left(\frac{\varepsilon_2-\varepsilon_3}{\varepsilon_2+\varepsilon_3}\right) \\ &+ \frac{3h\nu_\mathrm{e}}{8\sqrt{2}} \frac{(n_1^2-n_3^2)(n_2^2-n_3^2)}{(n_1^2+n_3^2)^{1/2}(n_2^2+n_3^2)^{1/2}[(n_1^2+n_3^2)^{1/2}+(n_2^2+n_3^2)^{1/2}]} \\ G_{132} &= -\frac{A_{132}}{12\pi d^2} \end{aligned}$$

where ε_i is the static dielectric constant of medium i, n_i is the refractive index of medium i, ν_e is the main electronic absorption frequency at about 3×10^{15} Hz, k is the Bolzmann constant, T is temperature, h is the Planck contant, and the subscript ν is the absorption frequency.

A_{total} is dominated by the dispersive term $A_{\nu>0}$, so we can use refractive indexes alone to predict the sign of the van der Waals interaction energy in our droplet-on-FLIPS systems. In addition, the calculations use the dielectric constants and refractive indices of the solvents of the ferrofluids or, in the case of 4 vol% ferrofluid, the properties of Krytox 100. This assumption is justified on the grounds that nanoparticles of iron oxides are shielded by fluorocarbon surfactants or polymers and that their van der Waals interactions with other media are therefore greatly reduced. The dominant role of dispersive forces is in the surface tension of pure ferrofluid: the dispersive components of the surface tension of ferrofluids 1 and 2 are estimated to be $13.7\,\mathrm{mN\,m^{-1}}$ and $12.4\,\mathrm{mN\,m^{-1}}$, respectively, using $\gamma_i^\mathrm{D} = A_{ii}/[24\pi(0.165\,\mathrm{nm})^2]$.

This method provides the same (correct) prediction for the formation of the wrapping layer as does the method in Extended Data Table 3. In essence, these calculations reflect the excess polarizability, or relative polarizability, of different phases: fluorocarbon is less polarizable than water or hydrocarbon, but more polarizable than air (vacuum)[23,38]. The physical constants of water, dodecane and air used are: $\varepsilon_\mathrm{water} = 80$, $n_\mathrm{water} = 1.333$; $\varepsilon_\mathrm{dodecane} = 2.01$, $n_\mathrm{dodecane} = 1.411$; and $\varepsilon_\mathrm{air} = 1$, $n_\mathrm{air} = 1$.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Method 3. We estimate the Hamaker constants A_{132} with relativistic retardation correction, the van der Waals interaction energy G_{132} and the disjoining pressure (results are presented in Extended Data Fig. 5). The calculations are based on the following equations, which are adapted from ref. [23]:

$$\varepsilon_i(i\xi_n) = 1 + \frac{{n_i}^2 - 1}{1 + (\xi_n/\omega_{\mathrm{UV}})^2}, \quad i = 1, 2, 3$$

$$\xi_n = \frac{2\pi kT}{\hbar} n, \quad n = 1, 2, 3, 4\ldots$$

$$R(l) = (1 + r_n)\mathrm{e}^{-r_n}$$

$$r_n(l) = \frac{2l}{c/[\varepsilon_3(i\xi_n)]^{1/2}}\xi_n$$

$$A_{132}(l) = \frac{3kT}{2}\sum_{\xi_n}\left[\frac{\varepsilon_1(i\xi_n) - \varepsilon_3(i\xi_n)}{\varepsilon_1(i\xi_n) + \varepsilon_3(i\xi_n)}\right]\left[\frac{\varepsilon_2(i\xi_n) - \varepsilon_3(i\xi_n)}{\varepsilon_2(i\xi_n) + \varepsilon_3(i\xi_n)}\right]R(l)$$

$$G_{132}(l) = -\frac{A_{132}(l)}{12\pi l^2}$$

$$p_{\mathrm{dis}} = -\frac{\partial G_{132}(l)}{\partial l}$$

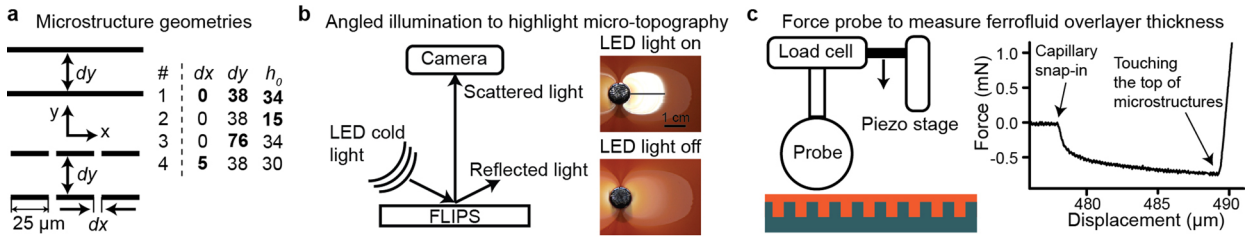
where, $\varepsilon_i(i\xi_n)$ is the dielectric response of medium i at imaginary frequency $i\xi_n$ and ξ_n are the sampling frequencies, or 'Matsubara frequencies'; $R(l)$ and $r_n(l)$ account for the retardation of van der Waals interactions over a distance l due to the finite velocity of electromagnetic waves; and $A_{132}(l)$ is the Hamaker constant for medium configuration 1–3–2, $G_{132}(l)$ is the corresponding van der Waals interaction energy and p_{dis} is the corresponding disjoining pressure.

This method is most useful when studing the thickness-dependent properties of wrapping layers, such as meta-stability, and can be extended to investigate the fine balances between disjoining, capillary and magnetic pressure.

Data availability. The datasets generated and analysed during this study are available from the corresponding author.

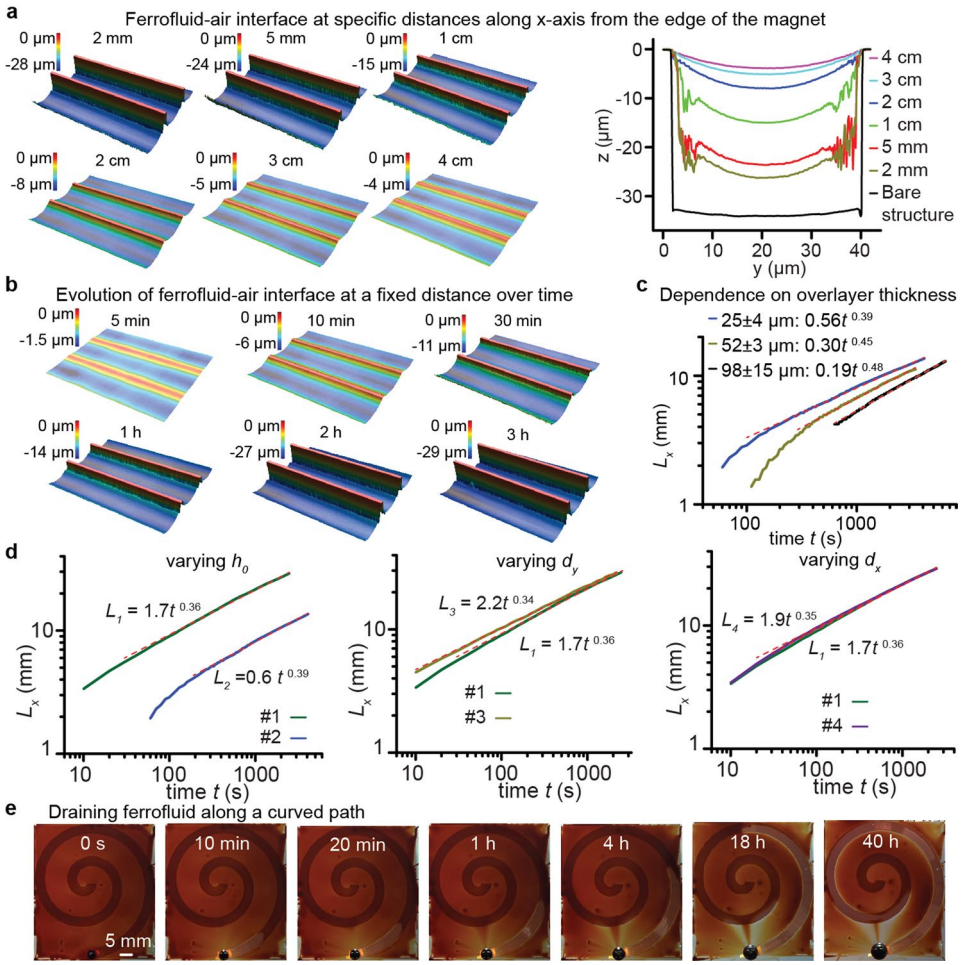
35. Pokroy, B., Epstein, A. K., Persson-Gulda, M. C. M. & Aizenberg, J. Fabrication of bioinspired actuated nanostructures with arbitrary geometry and stiffness. *Adv. Mater.* **21**, 463–469 (2009).
36. Kim, P., Adorno-Martinez, W. E., Khan, M. & Aizenberg, J. Enriching libraries of high-aspect-ratio micro- or nanostructures by rapid, low-cost, benchtop nanofabrication. *Nat. Protocols* **7**, 311–327 (2012).
37. Acheson, D. J. *Elementary Fluid Dynamics* Ch. 7 (Oxford Univ. Press, New York, 1990).
38. de Gennes, P.-G., Brochard-Wyart, F. & Quéré, D. *Capillarity and Wetting Phenomena: Drops, Bubbles, Pearls, Waves* Ch. 4 (Springer, Berlin, 2004).
39. Hardy, W. B. The tension of composite fluid surfaces and the mechanical stability of films of fluid. *Proc. R. Soc. Lond. A* **86**, 610–635 (1912).
40. Hardy, W. B. The tension of composite fluid surfaces.—No. II. *Proc. R. Soc. Lond. A* **88**, 313–333 (1913).
41. Smith, J. D. et al. Droplet mobility on lubricant-impregnated surfaces. *Soft Matter* **9**, 1772–1780 (2013).
42. Bonn, D., Eggers, J., Indekeu, J., Meunier, J. & Rolley, E. Wetting and spreading. *Rev. Mod. Phys.* **81**, 739–805 (2009).
43. Fowkes, F. M. Attractive forces at interfaces. *Ind. Eng. Chem.* **56**, 40–52 (1964).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



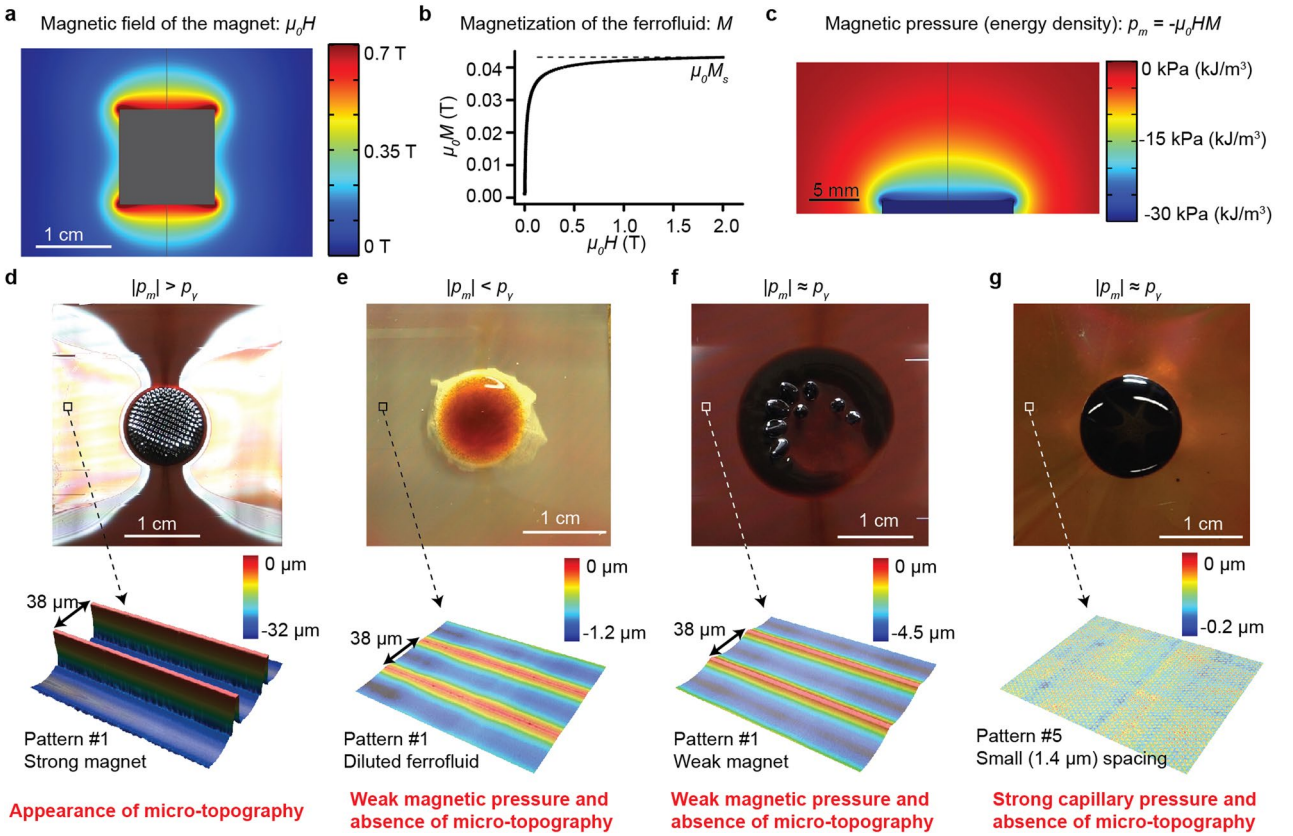
Extended Data Fig. 1 | Microstructure geometries and techniques for studying the dynamics of the topographical responses of FLIPS. **a**, Microstructure geometries used in the scaling analysis of the dynamics of micro-topographical response: arrays of microchannels (1–3) and an array of microplates (4). The comparisons between 1 and 2–4 show the influence of h_0, d_y and d_x, respectively. **b**, Angled illumination technique to highlight the micro-topographical region. A cold LED line light (Zeiss CL 6000 and line light S) illuminates the FLIPS sample from an incidence angle of about 30° to induce scattering from the micro-topographical region. The contrast due to the high brightness produced from the scattering facilitates the colour thresholding step in the video processing. The two images on the right show a comparison with and without the LED. **c**, Force-probe technique to measure the thickness of the ferrofluid overlayer. The diagram on the left is a schematic of the set-up. A spherical probe connected to a load cell is lowered slowly onto the FLIPS surface through the motion of a piezo stage. The speed of the approach (10 µm s^{-1}) and the sampling rate of the load cell (1 ms) determine the limit of the measurement resolution (10 nm). The plot on the right is a representative force–displacement curve from the measurement. The overlayer thickness is the distance between the capillary snap-in (the moment the probe touches the top of the overlayer) and the beginning of the steep increase in the force (the moment the probe touches the top of the microstructures).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



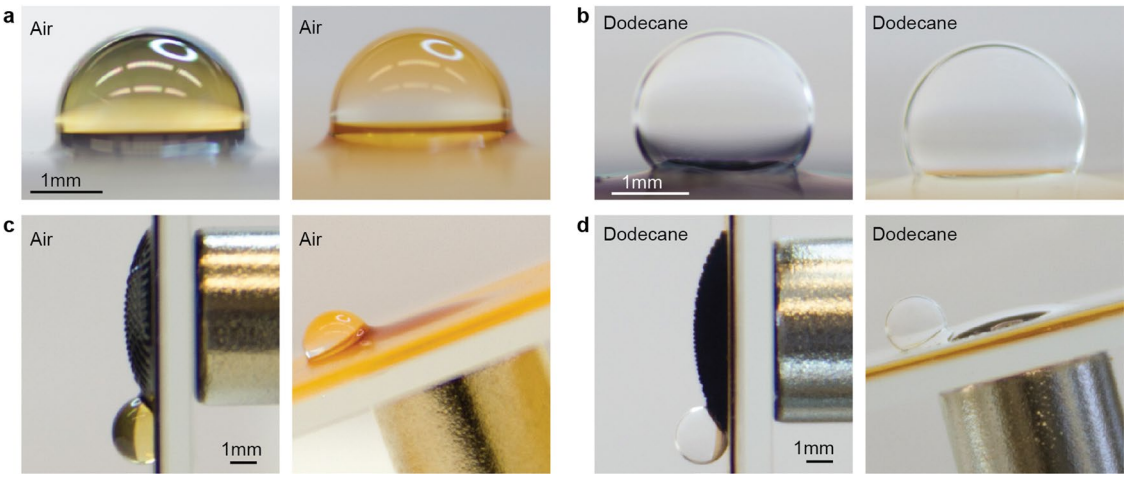
Extended Data Fig. 2 | Profiles of ferrofluid–air interfaces measured by a 3D laser-scanning confocal microscope, the dependence of the scaling relations on the overlayer thickness, and draining ferrofluid along a curved path. a, 3D profiles of the ferrofluid–air interface measured along the *x* axis at various distances from the edge of the magnet (labelled). The profiles show a gradual increase in the ferrofluid level inside the channel along the *x* axis, which corresponds to a gradual decrease in interfacial curvature. Measurements were performed on a 1 inch × 3 inch FLIPS sample after 2 h. Pattern 1 was used, and the channel direction was aligned to the long side of the FLIPS sample. On the right is a plot of the cross-sectional profiles of the ferrofluid–air interface. The signal is noisy in the high-curvature region near the edge, owing to the limit of the numerical aperture of the 100× long-working-distance objective used. **b**, 3D profiles of ferrofluid–air interface measured 1.5 cm away from the edge of the magnet over time (labelled). The FLIPS sample is the same as in **a**. The profiles show a gradual decrease in ferrofluid level inside the channel, which corresponds to a gradual increase in interfacial curvature. The corresponding plots of the cross-sectional profiles are presented in Fig. 1c. **c**, Dependence of the scaling relations on the overlayer thickness. Pattern 2 was used. Increasing the overlayer thickness decreases the prefactor but increases the power in the scaling relations. **d**, Effects of varying h_0, d_y and d_x on the scaling relations. The overlayer thicknesses are roughly 10–20 μm. **e**, Draining of ferrofluid along a curved path in a spiral pattern, demonstrating the ability of the porous-capillary flow to make turns.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



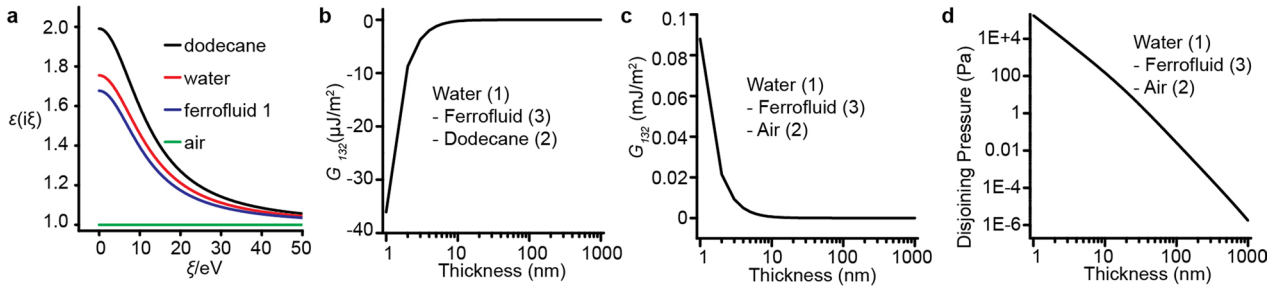
Extended Data Fig. 3 | Simulation of magnetic pressure and fine-tuning of the balance between magnetic and capillary pressure to create or conceal the micro-topography of FLIPS. a, Simulated magnetic field $B = \mu_0 H_0$ in free space surrounding a 0.5-inch-tall and 0.5-inch-diameter NdFeB magnet. **b**, Magnetization curve of ferrofluid 1 (from Ferrotec). The ferrofluid rapidly reaches its saturation magnetization $\mu_0 M_s \approx 0.04$ T. The saturation magnetization is regarded as a linear function of ferrofluid concentration in the following estimate. **c**, Magnetic pressure of ferrofluid 1 in the region above the magnet. The magnetic pressure (in kPa) can be considered as the magnetic energy density (in kJ m^{-3}). Ferrofluid flows from a high pressure (energy) region to a low pressure (energy) region. **d**, Turning on the micro-topographical response; $|p_m| \approx 10^4$ Pa $> 10^3$ Pa $\approx p_\gamma$. **e**–**g**, Turning off the micro-topographical response. **e**, Response when the ferrofluid is diluted to 4% by volume; $|p_m| \approx 10^2$ Pa $< 10^3$ Pa $\approx p_\gamma$. **f**, Response using a weaker magnet (Alnico); $|p_m| \approx 10^3$ Pa $\approx 10^3$ Pa $\approx p_\gamma$. **g**, The distance between the repeating units of the microstructure is reduced so that the capillary pressure increases; $|p_m| \approx 10^4$ Pa $\approx 10^4$ Pa $\approx p_\gamma$. The microstructure is pattern 5, which is a hexagonal array of posts of 1.5 μm in diameter and about 10 μm in height. The spacing between posts is 1.4 μm. The bottom panels in **d**–**g** show the corresponding 3D profiles of the ferrofluid–air interface near the magnet. Only in d is the interface pulled down towards the bottom of the microstructure.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



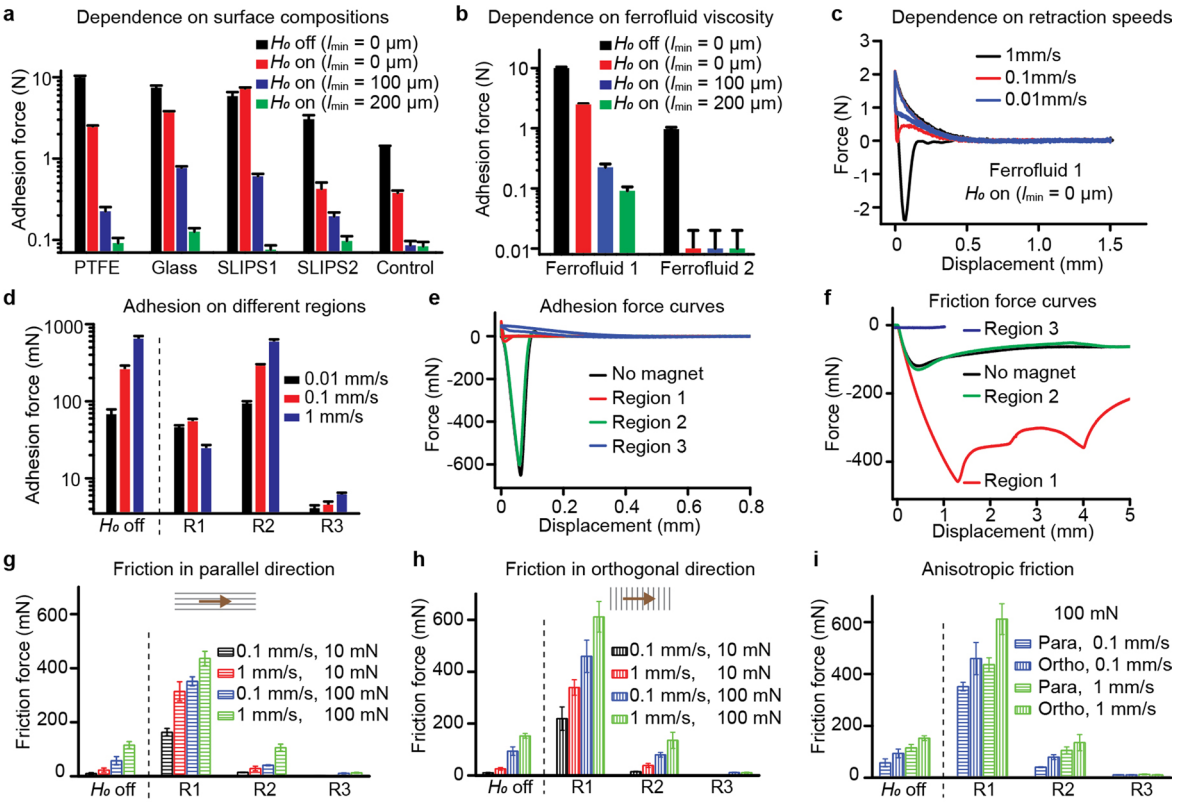
Extended Data Fig. 4 | Wrapping layers around a water droplet in air and in a hydrocarbon on FLIPS, and the ability of the macroscopic protuberance to hold a droplet at tilt angles. a, Photographs of a water droplet (about 5 μl) placed on a FLIPS with ferrofluid 1 (left) and diluted ferrofluid 1 (right). **b**, Photographs of a water droplet (about 5 μl) placed on a FLIPS with ferrofluid 1 (left) and diluted ferrofluid 1 (right) immersed in dodecane. **c**, Photographs of a water droplet (about 5 μl) held by a magnet at about 90° and 30° on a FLIPS with ferrofluid 1 (left) and diluted ferrofluid 1 (right), respectively. **d**, Photographs of a water droplet (about 5 μl) immersed in dodecane and held by a magnet at about 90° and 30° on a FLIPS with ferrofluid 1 (left) and diluted ferrofluid 1 (right), respectively. The comparison between **c** and **d** suggests that holding a droplet on the macroscopic protuberance does not require the existence of the wrapping layer and that capillary force is likely to dominate the interaction.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 5 | Calculations of the van der Waals interaction energy G_{132} and disjoining pressure. a, Dielectric response ε as a function of imaginary frequency ξ for the four phases in the droplet-on-FLIPS system. The ferrofluid line lies between air and water, but below both water and dodecane. **b**, Interaction energy as a function of the thickness of the ferrofluid wrapping layer for the water–ferrofluid–dodecane configuration. The energy decreases as the thickness decreases, indicating that the wrapping layer is unstable. **c**, Interaction energy as a function of the thickness of the ferrofluid wrapping layer for the water–ferrofluid–air configuration. The energy decreases as the wrapping layer thickness increases, indicating that the wrapping layer is stable. **d**, Disjoining pressure as a function of the thickness of the ferrofluid wrapping layer in the water–ferrofluid–air system. The pressures are positive, indicating the formation of a stable thin film or a stable wrapping layer.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 6 | Additional adhesion and friction data. **a**–**c**, Adhesion measured on FLIPS prepared with isotropic porous Teflon membrane substrates. l_{min} is the minimum distance between the test surface and the FLIPS, as shown in the schematics in Fig. 4a. **a**, Adhesion of FLIPS on a low-energy surface (PTFE), a high-energy surface (glass) and two composite surfaces: a fluorocarbon-oil-based SLIPS (SLIPS1) and a silicone-oil-based SLIPS (SLIPS2). The control is measured between two glass surfaces coated with ferrofluid only and without the porous membrane. **b**, Dependence of adhesion on ferrofluid viscosity. Ferrofluid 1 ($\mu = 0.367$ Pa s) is more viscous than ferrofluid 2 ($\mu = 0.040$ Pa s). **c**, Dependence of adhesion on retraction speed. The adhesion decreases with decreasing retraction speed. The test surface is a PTFE surface. Panels **b** and **c** suggest that viscous adhesion is the dominant adhesion mechanism. **d**–**i**, Adhesion and friction measurements on FLIPS with anisotropic substrates. Pattern 1 is used. The numbering of regions (R1, micro-topography; R2, flat region with a ferrofluid overlayer; R3, macro-topography) is the same as in Fig. 1. **d**, Adhesion without a magnet and on different regions with a magnet at three different retraction speeds. **e**, Sample adhesion force curves measured at 1 mm s^{-1}. **f**, Sample friction force curves measured at 1 mm s^{-1} and with 100-mN load. The value of the friction force is extracted from the lowest point along the curve. Note a signature of friction in the micro-topographical area of FLIPS, which is much higher than that on either flat or macro-topographical areas, characteristic of typical ferrofluid-coated non-structured surfaces. **g**, Friction forces without magnets and on different regions with a magnet in the direction parallel to the microchannels. Friction is highest in region 1, in the presence of micro-topography, and lowest in region 3, in the presence of macro-topography. The friction in region 2, with a flat fluid overlayer, is the same as for the entire FLIPS surface with no magnetic force. The difference can be as high as two orders of magnitude between regions 1 and 3, and one order of magnitude between regions 1 and 2 (or between region 1 and with no magnetic force). **h**, The corresponding friction forces in the orthogonal direction. **i**, Direct comparison of friction forces in the parallel and orthogonal directions. The error bars in all panels correspond to the standard deviations, calculated from a minimum of five measurements. For details, see Methods section 'Adhesion and friction experiments'.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Physical properties of fluorocarbon-based ferrofluids and perfluoropolyether oil (Krytox 100) provided by the manufacturers, and surface and interfacial tensions of liquids used here

Name	Density (g/cm^3)	Dielectric Constant ε	Refractive Index n	Dynamic Viscosity (Pa·s)	Kinematic Viscosity ($cm^2 \cdot s^{-1}$)	Volatility	Magnetic Saturation (Gauss)
FC Ferrofluid 1	1.8 - 2.0	2.0*	1.295*	0.367	1.9	0.5% at 100 °C after 3.5h	400
FC Ferrofluid 2	1.7 - 1.9	1.94*	1.28*	0.040	0.22	72.3% at 100 °C after 1.5h	400
Krytox 100	1.87	2.2	1.3	0.013	0.070	87% at 121 °C after 22h	-

Interfaces	Interfacial Tension (mN/m) ‡	Interfaces	Interfacial Tension (mN/m)
Water – air	73.4±0.2	FF 1 – water	42.9±0.7
Dodecane – air	25.3±0.1	FF 2 – water	37.4±0.6
FF 1 – air	17.1±0.2	4v% FF 1 – water	57.5±0.6
FF 2 – air	13.5±0.1	4v% FF 2 – water	55.3±0.8
4v% FF 1 – air†	16.0±0.4	FF 1 – dodecane	6.7±0.2
4v% FF 2 – air	15.6±0.5	FF 2 – dodecane	4.8±0.1
Dodecane – water	44.8±0.1#	4v% FF 1 – dodecane	6.6±0.2
		4v% FF 2 – dodecane	7.1±0.1

FC, fluorocarbon-based; FF, ferrofluid.
*Dielectric constants and refractive indices of pure solvents.
#The interfacial tension of a dodecane–water interface is highly sensitive to trace amounts of impurities: the value listed is for dodecane of reagent-plus grade from Aldrich; other measured values include $52.2 \pm 0.8\,mN\,m^{-1}$ for analytical-standard grade from Aldrich and $45.7 \pm 1.0\,mN\,m^{-1}$ for 99+% grade from Alpha Aesar.
†v% represents percentage by volume.
‡The errors correspond to the standard deviations, calculated from at least six measurements. For details, please see Methods section 'Control of the wrapping layer formation and droplet mixing'.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 2 | Initial spreading coefficients

Interface 1-2 (γ_{12})	Interface 3-1 (γ_{31})	Interface 3-2 (γ_{32})	Initial Spreading Coefficient $S_i = \gamma_{12} - (\gamma_{31} + \gamma_{32})$ (mN/m) *	Agreement with Experiments
Air – water	FF 1 – air	FF 1 – water	13.4	Yes
Air – water	FF 2 – air	FF 2 – water	22.5	Yes
Air – water	4v% FF 1 – air	4v% FF 1 – water	-1.2	No
Air – water	4v% FF 2 – air	4v% FF 2 – water	2.5	Yes
Dodecane – water	FF 1 – dodecane	FF 1 – water	-4.8	Yes
Dodecane – water	FF 2 – dodecane	FF 2 – water	2.6	No
Dodecane – water	4v% FF 1 – dodecane	4v% FF 1 – water	-19.3	Yes
Dodecane – water	4v% FF 2 – dodecane	4v% FF 2 – water	-17.6	Yes

*$S_i > 0$ indicates spreading; $S_i < 0$ indicates no spreading.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 3 | The signs of Hamaker constants A_{132} and van der Waals interaction energy G_{132} estimated from combining relations

Medium 1	Medium 3	Medium 2	Hamaker Constant A_{132}	van der Waals interaction energy G_{132} *	Agreement with Experiments
Air	FF 1	Water	<0	>0 (Repulsive)	Yes
Air	FF 2	Water	<0	>0 (Repulsive)	Yes
Air	4v% FF 1	Water	<0	>0 (Repulsive)	Yes
Air	4v% FF 2	Water	<0	>0 (Repulsive)	Yes
Dodecane	FF 1	Water	>0	<0 (Attractive)	Yes
Dodecane	FF 2	Water	>0	<0 (Attractive)	Yes
Dodecane	4v% FF 1	Water	>0	<0 (Attractive)	Yes
Dodecane	4v% FF 1	Water	>0	<0 (Attractive)	Yes

*Only the sign of the van der Waals interaction energy is needed to predict the formation or absence of the wrapping layer.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 4 | Non-retarded Hamaker constants A_{132} and van der Waals interaction energy G_{132} based on Lifshitz theory

Medium 1 – 3 – 2	$A_{v=0}$ (J)	$A_{v>0}$ (J)	A_{Total} (J)	van der Waals interaction energy G_{132}	Agreement with Experiments
Air – FF1 – Water	-9.8×10^{-22}	-3.1×10^{-21}	-4.1×10^{-21}	>0 (Repulsive)	Yes
Air – FF2 – Water	-9.4×10^{-22}	-4.2×10^{-21}	-5.1×10^{-21}	>0 (Repulsive)	Yes
Air – 4v% FF1 – Water	-1.1×10^{-21}	-2.7×10^{-21}	-3.8×10^{-21}	>0 (Repulsive)	Yes
Air – 4v% FF2 – Water	-1.1×10^{-21}	-2.7×10^{-21}	-3.8×10^{-21}	>0 (Repulsive)	Yes
Dodecane – FF1 – Water	7.3×10^{-24}	1.2×10^{-21}	1.2×10^{-21}	<0 (Attractive)	Yes
Dodecane – FF2 – Water	5.2×10^{-23}	1.8×10^{-21}	1.9×10^{-21}	<0 (Attractive)	Yes
Dodecane – 4v% FF1 – Water	-1.3×10^{-22}	9.6×10^{-22}	8.3×10^{-22}	<0 (Attractive)	Yes

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

https://doi.org/10.1038/s41586-018-0234-8

Decarboxylative sp^3 C–N coupling via dual copper and photoredox catalysis

Yufan Liang[1,2], Xiaheng Zhang[1,2] & David W. C. MacMillan[1]*

Over the past three decades, considerable progress has been made in the development of methods to construct sp^2 carbon–nitrogen (C–N) bonds using palladium, copper or nickel catalysis[1,2]. However, the incorporation of alkyl substrates to form sp^3 C–N bonds remains one of the major challenges in the field of cross-coupling chemistry. Here we demonstrate that the synergistic combination of copper catalysis and photoredox catalysis can provide a general platform from which to address this challenge. This cross-coupling system uses naturally abundant alkyl carboxylic acids and commercially available nitrogen nucleophiles as coupling partners. It is applicable to a wide variety of primary, secondary and tertiary alkyl carboxylic acids (through iodonium activation), as well as a vast array of nitrogen nucleophiles: nitrogen heterocycles, amides, sulfonamides and anilines can undergo C–N coupling to provide *N*-alkyl products in good to excellent efficiency, at room temperature and on short timescales (five minutes to one hour). We demonstrate that this C–N coupling protocol proceeds with high regioselectivity using substrates that contain several amine groups, and can also be applied to complex drug molecules, enabling the rapid construction of molecular complexity and the late-stage functionalization of bioactive pharmaceuticals.

Within the field of organic chemistry, the efficient construction of C–N bonds is important owing to the prevalence of nitrogen-containing motifs in a wide array of natural products, pharmaceuticals and functional materials[3–5]. There are several notable routes to the formation of sp^2 C–N bonds, including the Buchwald–Hartwig reaction[1], Ullmann coupling[6] and Chan–Lam amination[7]. However, sp^3 C–N bond formation typically relies on classical methods, such as nucleophilic substitution reactions between nitrogen nucleophiles and alkyl halides[8], Mitsunobu alkylations of alcohols using nitrogen nucleophiles[9], reductive amination with carbonyls[10] or olefin hydroamination[11]. Recently, several research groups have reported transition-metal-catalysed variants of the alkylation reaction of nitrogen nucleophiles with aliphatic halides[12,13]. In 1894, Curtius reported the rearrangement of acyl azides to form C–N-containing aliphatic substrates[14]. We questioned whether it might be possible to expand this concept to complex, medicinally relevant nitrogen-bearing fragments, thereby accelerating access to drug-like complexity in one step.

The synergistic merger of photoredox[15] and transition metal catalysis (termed metallaphotoredox catalysis) has resulted in the development of many cross-coupling reactions that are now being widely adopted within the pharmaceutical sector[16]. The combination of nickel and photoredox catalysis has enabled the efficient construction of C(sp^3)–C(sp^2) and C(sp^3)–C(sp^3) bonds using abundant alkyl carboxylic acids[17] and alcohols[18]. Here we show that metallaphotoredox catalysis involving copper in place of nickel enables alkyl sp^3 C–N bond formation in a generic sense without the use of alkyl halides or other prototypical electrophiles. More specifically, we hoped to merge the capacity of photoredox reactions to form alkyl radicals from iodonium carboxylates (derived in situ from carboxylic acids) with the long-established propensity of copper to participate in reductive elimination to form carbon–heteroatom bonds[2]. By taking advantage of the widely abundant nature of alkyl carboxylic acids and nitrogen nucleophiles such as heteroaromatics, sulfonamides, amides and anilines, we hoped to devise a new fragment coupling reaction that would be broadly useful yet mechanistically orthogonal to established alkylation reactions (Fig. 1). Recently, two methods have been reported with this aim in mind[19,20], and these illustrate the capacity of copper to function in decarboxylative mechanisms.

A detailed mechanism for the proposed decarboxylative coupling of sp^3 carbon with nitrogen nucleophiles is outlined in Fig. 2a. Excitation of photocatalyst Ir(F-Meppy)$_2$(dtbbpy)PF$_6$ (**1**) (F-Meppy = 2-(4-fluorophenyl)-5-(methyl)pyridine, dtbbpy = 4,4′-di-*tert*-butyl-2,2′-bipyridine) is known to generate the long-lived triplet-excited-state *Ir^{III} complex **2** (with a lifetime, τ, of 1.1 μs)[21]. At the same time, we proposed that coordination of the nitrogen nucleophile **11** with a copper(I) precatalyst followed by deprotonation would readily form the copper(I)-amido species **3**. The excited state *Ir^{III} complex **2** ($E_{1/2}^{red}$ [*Ir^{III}/Ir^{II}] = 0.94 V versus the standard calomel electrode (SCE) in CH_3CN)[21] should rapidly oxidize this copper(I) complex **3** ($E_{1/2}^{red}$ [Cu^{II}(BPhen)$_2$/Cu^{I}(BPhen)$_2$] = 0.08 V versus SCE in DMF, BPhen = 4,7-diphenyl-1,10-phenanthroline)[22] to generate the corresponding copper(II)-amido system **4** and the corresponding iridium(II) complex **5**. At this stage we considered that iodomesitylene dicarboxylate **8** (which is preformed via the mixing of carboxylic acid **6** and iodomesitylene diacetate **7**, see Supplementary Information) would be readily reduced by the newly formed iridium(II) species **5** ($E_{1/2}^{red}$ [Ir^{III}/Ir^{II}] = −1.50 V versus SCE in CH_3CN[21], E_p [**8**/**8**$^{\bullet-}$] = −1.14 V versus SCE in CH_3CN) to generate a carboxyl radical, which upon CO_2 extrusion[23,24] would produce the desired alkyl radical **9**, while reconstituting the ground-state photocatalyst **1**. At this stage, we anticipated that copper(II)-amido complex **4** would capture alkyl radical **9** to form copper(III) complex **10**, which upon reductive elimination[25] would yield the desired fragment-coupled sp^3 C–N bearing adduct **12** and regenerate copper(I) catalyst **3**.

We first examined the proposed sp^3 C–N coupling using three electronically disparate nitrogen nucleophiles (indole **11a**, azaindole **11b** and indazole **11c**, Fig. 2b), along with cyclohexyl carboxylic acid **6** as the alkylating reagent, and a wide range of copper(I) and photoredox catalysts. The desired decarboxylative sp^3 C–N coupling can be achieved in good to excellent efficiency for all three substrates (60%, 76% and 90% yield, respectively) using Ir(F-Meppy)$_2$(dtbbpy)PF$_6$ as the photocatalyst, CuTC (TC = thiophene-2-carboxylate) as the copper catalyst, BPhen or dOMe-Phen (dOMe-Phen = 4,7-dimethoxy-1,10-phenanthroline) as the ligand, BTMG (BTMG = 2-*tert*-butyl-1,1,3,3-tetramethylguanidine) as the base, with exposure to 34-W blue light-emitting diodes (LEDs). Notably, a series of control experiments revealed that although the copper(I) catalyst is essential for the desired C–N bond formation in all cases, the absence of light and/or photocatalyst has a profound effect on reaction efficiency depending on the nitrogen nucleophile used. More specifically, the alkylation of indazole **11c** is successful with or without photocatalysis (90% compared with 86% yield, respectively), whereas indole **11a** and azaindole **11b** achieve markedly improved yields and reaction times when light

[1]Merck Center for Catalysis at Princeton University, Princeton, NJ, USA. [2]These authors contributed equally: Yufan Liang, Xiaheng Zhang. *e-mail: dmacmill@princeton.edu

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

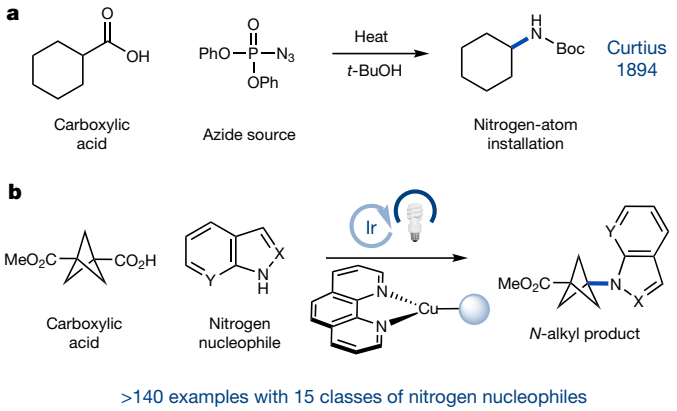
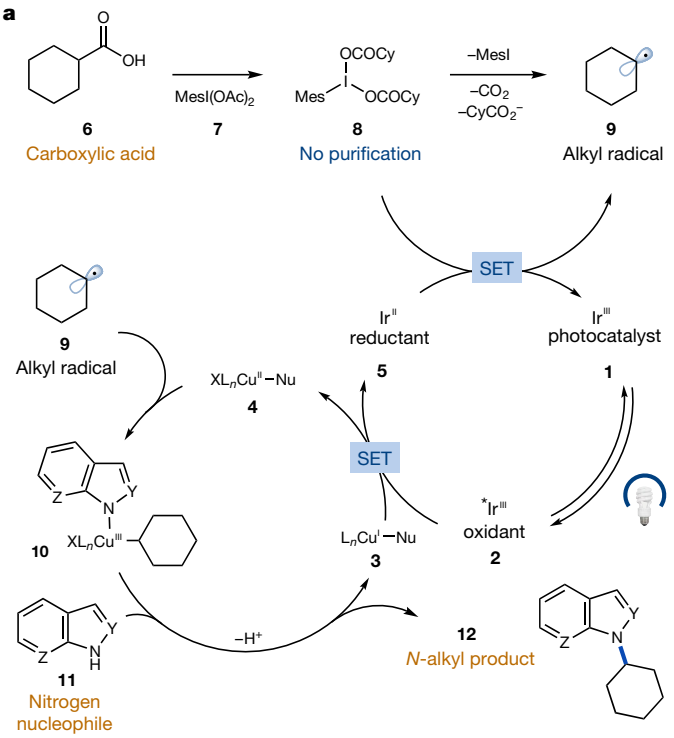


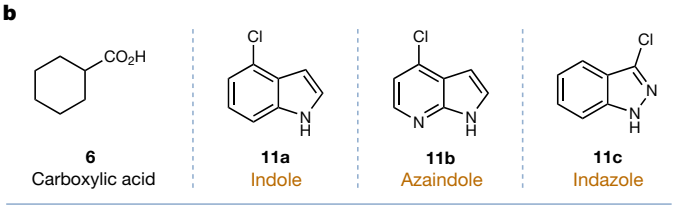
Fig. 1 | Decarboxylative nitrogen-nucleophile fragment coupling. **a**, An example of the Curtius[14] rearrangement of acyl azides to form C–N-containing aliphatic substrates. **b**, A general platform for decarboxylative sp^3 C–N coupling can be realized by the combination of copper catalysis and photoredox catalysis. A broad range of readily available carboxylic acids and nitrogen nucleophiles are used to generate various *N*-alkyl products. Boc, *tert*-butoxycarbonyl; Ph, phenyl; X and Y, carbon or nitrogen atom.

and the iridium photocatalyst are combined with copper (indole **11a**, 60% compared with 7% yield; azaindole **11b**, 76% compared with 47% yield). We speculate that a non-photonic mechanism is possible when copper(I)-amido species **3** is sufficiently electron-rich to undergo direct electron-transfer with the iodomesitylene dicarboxylate **8**, thereby removing the requirement for a redox catalyst to act as an electron shuttle. However, when copper(I)-amido species **3** is not sufficiently reducing, or is formed slowly, the presence of a photoexcited electron-shuttle catalyst becomes essential to achieve useful efficiencies. Indeed, this latter case was found to be the most common, with non-photonic conditions leading to useful yields with only a small subset of substrates (yields for non-photonic conditions are reported in parentheses in Figs. 3 and 4). As such, the combination of copper and photoredox catalysts with light was identified as the superior protocol with which to evaluate a broad range of sp^3 C–N cross-coupling reactions.

Having established the preferred reaction conditions, we next examined the generality of this new C–N fragment coupling by exploring the scope of the carboxylic-acid alkylation partner (Fig. 3). Notably, a diverse range of alkyl carboxylic acids can be used in this new protocol to deliver the *N*-alkyl heteroaryl derivatives in good to excellent efficiency. It is important to highlight that in all cases the reactions were complete at room temperature within 1 h. Another feature of this protocol is that only one regioisomer of the product is produced for all cases shown in Fig. 3, which offers a notable advantage when compared to traditional *N*-alkylation reactions. Indeed, a broad series of differentially substituted primary alkyl acids can readily participate in this new C–N coupling (**13**–**21**, 50–82% yield). Moreover, these mild reaction conditions are compatible with a range of common functional groups, such as terminal olefins (**17**, 64% yield), terminal alkynes (**18**, 80% yield), nitro groups (**19**, 50% yield), esters (**21**, 80% yield) and protected amines (**16** and **20**, 63% and 82% yield, respectively). In contrast to many established alkylation reactions, steric encumbrance proximal to the acid group is also well tolerated, as exemplified by the successful incorporation of a neopentyl system (**15**, 54% yield). Moreover, direct *N*-methylation (**13**, 65% yield) can also be realized by carrying out the C–N coupling protocol using commercial $MesI(OAc)_2$ **7**. Finally, we successfully applied this coupling technology to three complex natural products that contain alcohols, ketones and internal alkenes. All were found to participate in decarboxylative *N*-alkylation with high efficiency (**22**–**24**, 54%–76% yield).

We next sought to examine the scope of secondary alkyl carboxylic acids as alkylating agents. Importantly, a large array of ring-bearing carboxylates can be used to access *N*-cycloalkylation adducts in good





Conditions	Yield	Yield	Yield
No Cu(I)	<1%	<1%	<1%
With Cu(I), no light	7%	47%	86%
With Cu(I), with light	11%	50%	86%
With Cu(I), with light, with photocatalyst	60%	76%	90%

Fig. 2 | Catalytic cycles and control experiments. **a**, A proposed mechanism is outlined. Photocatalyst **1** is excited by visible light to produce a long-lived triplet excited state (**2**), which can readily oxidize copper(I) catalyst **3** to yield copper(II) species **4**. The reduced iridium(II) complex **5** can then be oxidized by iodomesitylene dicarboxylate **8**, which is derived from the reaction of carboxylic acid **6** and iodomesitylene diacetate **7**, to release alkyl radical **9** and photocatalyst **1**. Concurrently in the copper catalytic cycle, copper(II)-amido complex **4** can capture alkyl radical **9** to form the highly unstable copper(III) complex **10**. Reductive elimination from complex **10** provides the desired sp^3 C–N coupled product and regenerates the copper(I) catalyst **3** after coordination with the nitrogen nucleophile **11**. **b**, Control experiments were performed with three different nitrogen nucleophiles. In all cases, the presence of both the copper(I) catalyst and the photoredox catalyst under light irradiation conditions is crucial to achieve the best efficiency for the C–N coupling reactions. Cy, cyclohexyl; L, ligand; Mes, mesityl; Nu, nucleophile; SET, single-electron transfer; X, anionic ligand, such as a carboxylate; Y and Z, carbon or nitrogen atoms.

to excellent efficiency (**26**–**31**, 40%–85% yield). Moreover, this new transformation is not limited to cyclic systems, as exemplified by the rapid incorporation of acyclic secondary alkyl groups (**25**, 61% yield), a transformation that is not readily achieved using established alkylation protocols. An additional six examples using other secondary alkyl acids for this sp^3 C–N coupling are detailed in Supplementary Information and Extended Data Fig. 1.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

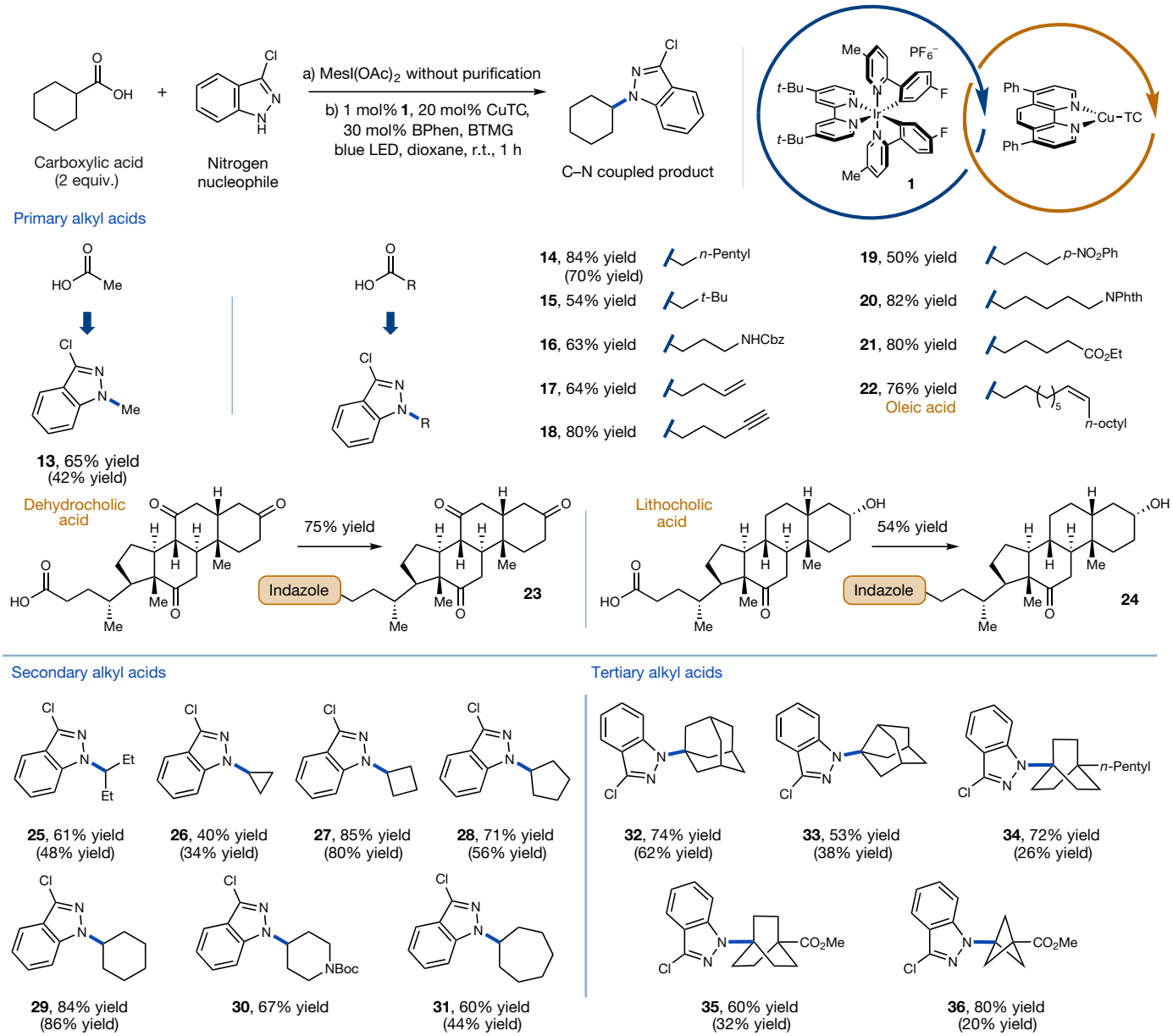


Fig. 3 | Decarboxylative C–N couplings of 3-chloroindazole with a range of alkyl carboxylic acids. A wide variety of alkyl carboxylic acids can be cross-coupled with 3-chloroindazole. The carboxylic acid is added as part of step a, and the nitrogen nucleophile as part of step b. The protocol provides the product as a single regioisomer in all cases. The yields shown first are isolated yields of reactions conducted according to our standard protocol; yields shown in parentheses are for reactions that were conducted in the absence of light and a photocatalyst and were determined by ^{1}H nuclear magnetic resonance (^{1}H NMR) spectroscopy with an internal standard. See Supplementary Information for full experimental details, and Extended Data Fig. 1 for additional examples. All reactions were replicated at least twice for consistency. Ac, acetyl; BTMG, 2-tert-butyl-1,1,3,3-tetramethylguanidine; Cbz, carboxybenzyl; Et, ethyl; Me, methyl; NPhth, phthalimide; TC, thiophene-2-carboxylate; r.t., room temperature.

A notable feature of this decarboxylative C–N coupling method is the number of tertiary-carbon-bearing carboxylic acids that can be readily used to prepare heteroaryl *N*-tertiary alkyl derivatives with good to excellent efficiency (**32**–**36**, 53%–80% yield). This is especially important given that the *N*-alkylation using tertiary alkyl halides can often be elusive, if not impossible, using traditional technologies. Given the recent interest in the incorporation of rigid bicyclic structures into drug-like compounds, such as the bicyclo[1.1.1]pentane core (as shown in product **36**)[26], we expect this light-mediated *N*-alkylation protocol to be immediately applicable to a large range of medicinal chemistry programs. Indeed, the capacity to access these bridged bicyclic aryl isosteres in only one operation using commercial carboxylic acids and iodomesitylene diacetate (at room temperature in less than 1 h) should enable the rapid adoption of this method.

Next, we turned our attention to the scope of the nitrogen-nucleophile component in this new catalytic alkylation protocol (Fig. 4). Almost every class of medicinally relevant nitrogen heterocycle, including, but not limited to, indazoles (**37** and **38**, 81% and 73% yield, respectively), azaindoles (**39** and **40**, 89% and 75% yield, respectively), indoles (**41**, 58% yield), pyrazoles (**42** and **43**, 98% and 68% yield, respectively), pyrroles (**51**, 75% yield), imidazoles (**52**, 68% yield), triazoles (**53**, 90% yield), benzimidazoles (**54**, 67% yield), benzotriazoles (**55**, 80% yield), purines (**56**, 60% yield) and carbazoles (**57**, 46% yield), can be successfully used to deliver *N*-alkyl products in good to excellent efficiency. Moreover, an additional 42 examples using other nitrogen heterocycles are detailed in Supplementary Information and Extended Data Figs. 2 and 3. For nucleophiles that are as acidic as or more acidic than pyrazoles (such as indazoles, triazoles and imidazoles), the addition of an exogenous base is not generally necessary. The carboxylate anion, which is generated

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

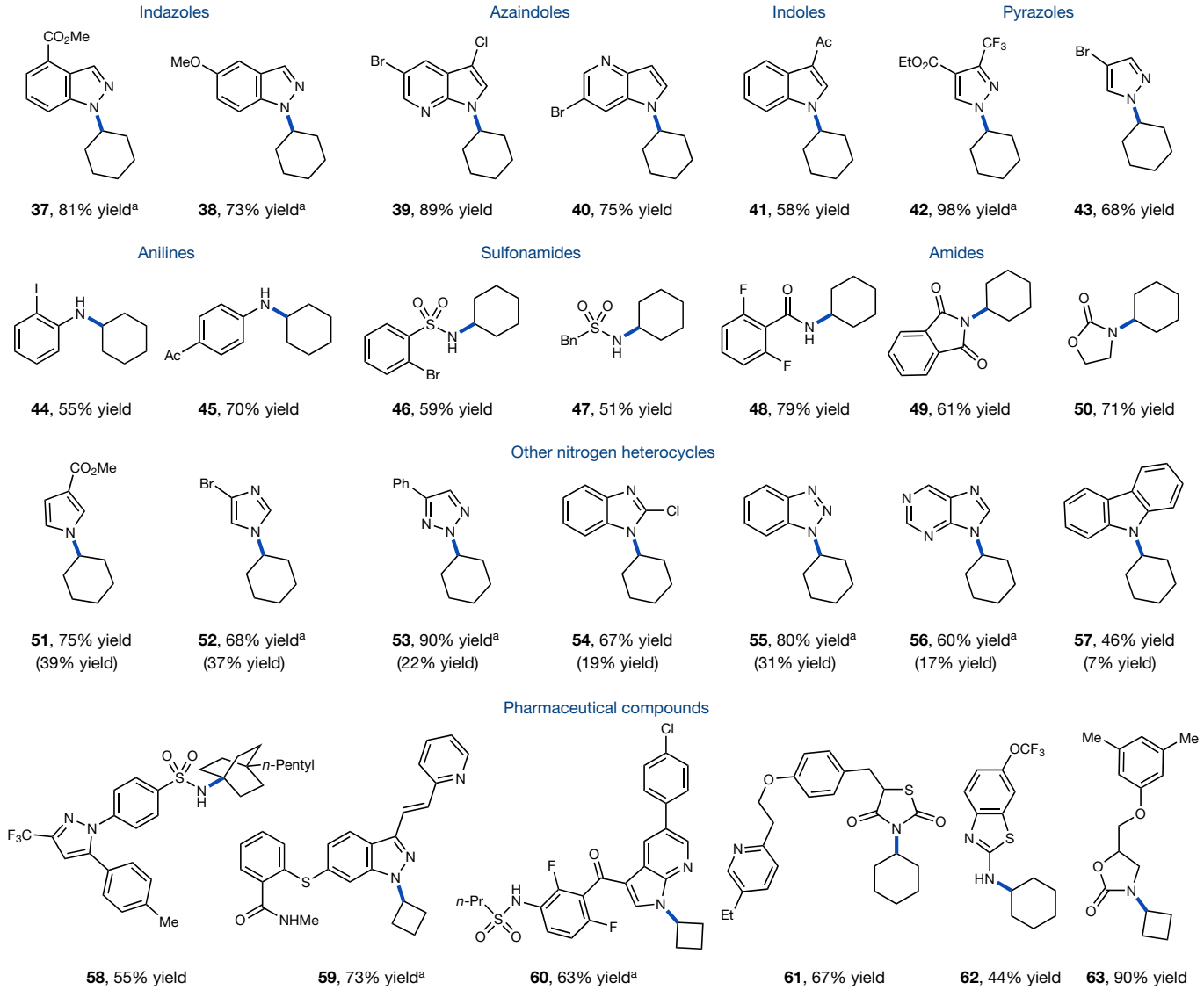


Fig. 4 | Decarboxylative C–N couplings of cyclohexyl carboxylic acid with various nitrogen nucleophiles. This new C–N bond-forming protocol shows a markedly broad scope with respect to the nitrogen nucleophiles. Almost every class of important nitrogen heterocycle can provide *N*-alkylated products in good yields and excellent regioselectivity. Less nucleophilic substrates and complex drug molecules are all viable coupling partners. The yields shown first are isolated yields for the decarboxylative C–N coupling step; yields shown in parentheses are for reactions that were conducted in the absence of light and a photocatalyst and were determined by ^{1}H NMR with an internal standard. See Supplementary Information for full experimental details, and Extended Data Figs. 2–4 for additional examples. [a]Single regioisomer. Bn, benzyl; Pr, propyl.

upon the reduction of iodomesitylene dicarboxylate **8** (Fig. 2a), can function as a weak base. One notable feature is that this heterocycle C–N forming mechanism exhibits excellent regioselectivity (Fig. 4) with substrates that possess several *N*-alkylation sites (for example, pyrazoles, imidazoles, triazoles, indazoles and benzimidazoles). At this stage, we presume that the considerable steric bulk of the ligated copper complex ensures that the sp^3 C–N coupling takes place at the least hindered site of these heteroaromatic nucleophiles. This outcome is in sharp contrast to classical alkylation methods that typically lead to regioisomeric mixtures. Several of these nitrogen heterocycles were also tested using our non-photonic conditions, and the corresponding yields are shown in parentheses in Fig. 4. It is clear that although C–N formation can be achieved with these nucleophiles in the absence of light (**51**–**57**, 7–39% yield), the dual copper and photoredox protocol enables a more extensive scope and substantially higher efficiencies across the board, providing a more general protocol for this new C–N heterocyclic coupling reaction.

As shown in Fig. 4, this decarboxylative C–N coupling method is not limited to the cross-coupling of nitrogen heterocycles. Under our optimized conditions, a large selection of electron-deficient (and also less acidic) nitrogen nucleophiles, including anilines (**44** and **45**, 55% and 70%, respectively), aryl sulfonamides (**46**, 59% yield), alkyl sulfonamides (**47**, 51% yield), aryl amides (**48**, 79% yield), phthalimides (**49**, 61% yield) and cyclic carbamates (**50**, 71% yield), were found to participate readily in this sp^3 C–N coupling. Notably, functional groups including aryl iodides (**44**, 55% yield), aryl bromides (**46**, 59% yield) and ketones (**45**, 70% yield) were readily tolerated, a useful feature with respect to further synthetic manipulation.

A long-established problem in the functionalization of primary amines with alkyl halides is the formation of polyalkylation products, given that the initial secondary amine adduct is more nucleophilic than the starting amine. Notably, as shown in Fig. 4, only monoalkylated products are obtained when primary amides, sulfonamides and anilines are used in this new copper-catalysed protocol (**44**–**48**, also see Supplementary Information for additional examples using primary alkyl acids). Moreover, an additional 27 examples using other electron-deficient nitrogen nucleophiles are detailed in Supplementary Information and Extended Data Fig. 3. As already highlighted, this

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

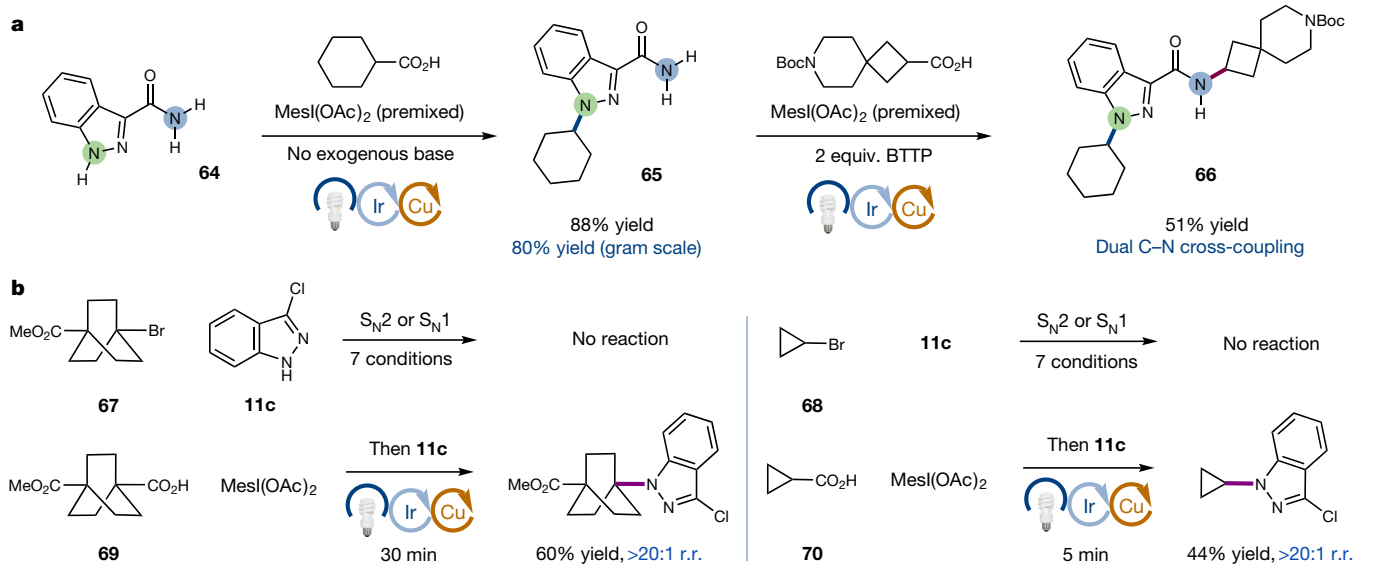


Fig. 5 | Sequential C–N couplings and comparisons with nucleophilic substitutions. a, Sequential decarboxylative C–N couplings can be realized using indazole derivative **64** that contains two nucleophilic sites, demonstrating the construction of molecular complexity. The *N*,*N*′-dialkylated product, which contains two different alkyl groups, can be easily generated through two C–N coupling reactions using different alkyl acids under different reaction conditions. All yields in this section are isolated yields for the decarboxylative C–N coupling step. **b**, Comparing the current decarboxylative C–N coupling protocol with classical nucleophilic substitution methods using two alkyl electrophiles demonstrates the complementary nature of this new method. All yields in this section were determined by ^{1}H NMR studies with an internal standard. See Supplementary Information for full experimental details and additional examples. BTTP, *tert*-butylimino-tri(pyrrolidino)phosphorane; r.r., regiomeric ratio.

coupling protocol does not appear to be negatively influenced by steric constraints, as *ortho*-substituted and *ortho*-,*ortho*-disubstituted anilines, sulfonamides and amides can be used readily (**44**, **46**, and **48**, 55%–79% yield).

To illustrate the utility of this new transformation with respect to drug discovery, we examined this decarboxylative sp^3 C–N coupling using six known pharmaceuticals (Celebrex, Axitinib, Zelboraf, Actos, Rilutek and Skelaxin). Using three separate carboxylic acids, we were able to achieve decarboxylative alkylation in all cases in good to excellent yields (**58**–**63**, 44%–90% yield). An additional ten examples of pharmaceutical functionalization using this technology are described in Supplementary Information and Extended Data Fig. 4.

For substrates with several nucleophilic sites, achieving regioselective monofunctionalization has been a long-standing challenge[27]. Therefore, we were pleased to find that this new alkylation technology can be applied sequentially to the same drug molecule to achieve selective alkylation at two discrete nitrogen positions through the judicious choice of reaction conditions. As demonstrated in Fig. 5a, heterocycle **64** contains both an indazole nitrogen and a primary amide; however, when the coupling protocol is performed without an exogenous base, regioselective *N*-alkylation of the indazole was observed in 80% yield. Moreover, we have carried out this regioselective *N*-alkylation step on a 7.4-mmol scale to prepare 1.45 g of the *N*-cyclohexyl indazole derivative **65**. The origins of regioselectivity in this decarboxylative coupling arise from the relative acidity of the two N–H moieties in substrate **64** (that is, indazole and amide). More specifically, when no exogenous base is used, the carboxylate anion (formed by reduction of the iodomesitylene dicarboxylate **8**) can function as a weak base and thereby selectively deprotonate the more acidic indazole nitrogen (pK_a (indazole) <19.8 in DMSO[28,29]; pK_a (phenyl acetamide) $=23.3$ in DMSO[28]) upon coordination to the copper catalyst, which in turn leads to regioselective *N*-alkylation of the indazole N1 position. Perhaps most important, subsequent exposure of the resulting *N*-cyclohexyl indazole **65** to our coupling protocol with a second carboxylic acid in the presence of a strong organic base, BTTP (BTTP = *tert*-butyliminotri(pyrrolidino) phosphorane) leads to a second *N*-alkylation on the remaining amide nitrogen, again with a useful yield. We anticipate that the capacity to perform regioselective and sequential C–N coupling steps as a function of relative N–H acidities will have substantial benefit with respect to the step economy of building complex molecules, by removing the need to install and remove nitrogen protecting groups.

Finally, to further showcase the utility of this new sp^3 C–N coupling method, we performed a series of experiments to compare the decarboxylative protocol with traditional nucleophilic substitution reactions (Fig. 5b). Under various classical S_N2 and S_N1 alkylation conditions, tertiary alkyl bromide **67** and bromocyclopropane (**68**) failed to react with 3-chloroindazole to generate the desired *N*-alkyl products, which is consistent with a lack of literature precedent for using these alkyl bromides with indazole nucleophiles. By contrast, through decarboxylative C–N couplings using commercially available carboxylic acids **69** and **70** and our mild catalytic protocol, the desired *N*-alkyl products can be obtained in good yields and excellent regioselectivities within 30 min at room temperature using an integrated photoreactor[30]. As such, we anticipate that this new decarboxylative coupling strategy will provide a useful, complementary new approach to sp^3 C–N alkylation.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0234-8.

Received: 25 January 2018; Accepted: 13 April 2018;
Published online: 20 June 2018

1. Ruiz-Castillo, P. & Buchwald, S. L. Applications of palladium-catalyzed C–N cross-coupling reactions. *Chem. Rev.* **116**, 12564–12649 (2016).
2. Bhunia, S., Pawar, G. G., Kumar, S. V., Jiang, Y. & Ma, D. Selected copper-based reactions for C–N, C–O, C–S, and C–C bond formation. *Angew. Chem. Int. Ed.* **56**, 16136–16179 (2017).
3. Knölker, H.-J. (ed.) *The Alkaloids: Chemistry and Biology* Vol. 70 (Elsevier, San Diego, 2011).
4. Vitaku, E., Smith, D. T. & Njardarson, J. T. Analysis of the structural diversity, substitution patterns, and frequency of nitrogen heterocycles among U.S. FDA approved pharmaceuticals. *J. Med. Chem.* **57**, 10257–10274 (2014).
5. Ćirić-Marjanović, G. Recent advances in polyaniline research: polymerization mechanisms, structural aspects, properties and applications. *Synth. Met.* **177**, 1–47 (2013).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

6. Sambiagio, C., Marsden, S. P., Blacker, A. J. & McGowan, P. C. Copper catalysed Ullmann type chemistry: from mechanistic aspects to modern development. *Chem. Soc. Rev.* **43**, 3525–3550 (2014).
7. Qiao, J. X. & Lam, P. Y. S. Copper-promoted carbon–heteroatom bond cross-coupling with boronic acids and derivatives. *Synthesis* **2011**, 829–856 (2011).
8. Salvatore, R. N., Yoon, C. H. & Jung, K. W. Synthesis of secondary amines. *Tetrahedron* **57**, 7785–7811 (2001).
9. Swamy, K. C. K., Kumar, N. N. B., Balaraman, E. & Kumar, K. V. P. P. Mitsunobu and related reactions: advances and applications. *Chem. Rev.* **109**, 2551–2651 (2009).
10. Abdel-Magid, A. F. & Mehrman, S. J. A review on the use of sodium triacetoxyborohydride in the reductive amination of ketones and aldehydes. *Org. Process Res. Dev.* **10**, 971–1031 (2006).
11. Huang, L., Arndt, M., Gooßen, K., Heydt, H. & Gooßen, L. J. Late transition metal-catalyzed hydroamination and hydroamidation. *Chem. Rev.* **115**, 2596–2697 (2015).
12. Matier, C. D., Schwaben, J., Peters, J. C. & Fu, G. C. Copper-catalyzed alkylation of aliphatic amines induced by visible light. *J. Am. Chem. Soc.* **139**, 17707–17710 (2017).
13. Peacock, D. M., Roos, C. B. & Hartwig, J. F. Palladium-catalyzed cross coupling of secondary and tertiary alkyl bromides with a nitrogen nucleophile. *ACS Cent. Sci.* **2**, 647–652 (2016).
14. Curtius, T. 20. Hydrazide und azide organischer säuren I. Abhandlung. *J. Prakt. Chem.* **50**, 275–294 (1894).
15. Shaw, M. H., Twilton, J. & MacMillan, D. W. C. Photoredox catalysis in organic chemistry. *J. Org. Chem.* **81**, 6898–6926 (2016).
16. Twilton, J., Le, C., Zhang, P., Shaw, M. H., Evans, R. W. & MacMillan, D. W. C. The merger of transition metal and photocatalysis. *Nat. Rev. Chem.* **1**, 0052 (2017).
17. Zuo, Z. et al. Merging photoredox with nickel catalysis: coupling of α-carboxyl sp^3-carbons with aryl halides. *Science* **345**, 437–440 (2014).
18. Zhang, X. & MacMillan, D. W. C. Alcohols as latent coupling fragments for metallaphotoredox catalysis: sp^3–sp^2 cross-coupling of oxalates with aryl halides. *J. Am. Chem. Soc.* **138**, 13862–13865 (2016).
19. Zhao, W., Wurz, R. P., Peters, J. C. & Fu, G. C. Photoinduced, copper-catalyzed decarboxylative C–N coupling to generate protected amines: an alternative to the Curtius rearrangement. *J. Am. Chem. Soc.* **139**, 12153–12156 (2017).
20. Mao, R., Frey, A., Balon, J. & Hu, X. Decarboxylative C(sp^3)–N cross-coupling via synergistic photoredox and copper catalysis. *Nat. Catal.* **1**, 120–126 (2018).
21. Lowry, M. S. et al. Single-layer electroluminescent devices and photoinduced hydrogen production from an ionic iridium(III) complex. *Chem. Mater.* **17**, 5712–5719 (2005).
22. Sanna, G., Pilo, M. I., Zoroddu, M. A. & Seeber, R. Electrochemical and spectroelectrochemical study of copper complexes with 1,10-phenanthrolines. *Inorg. Chim. Acta* **208**, 153–158 (1993).
23. He, Z., Bae, M., Wu, J. & Jamison, T. F. Synthesis of highly functionalized polycyclic quinoxaline derivatives using visible-light photoredox catalysis. *Angew. Chem. Int. Ed.* **53**, 14451–14455 (2014).
24. Minisci, F., Vismara, E., Fontana, F. & Barbosa, M. C. N. A new general method of homolytic alkylation of protonated heteroaromatic bases by carboxylic acids and iodosobenzene diacetate. *Tetrahedr. Lett.* **30**, 4569–4572 (1989).
25. Tran, B. L., Li, B., Driess, M. & Hartwig, J. F. Copper-catalyzed intermolecular amidation and imidation of unactivated alkanes. *J. Am. Chem. Soc.* **136**, 2555–2563 (2014).
26. Stepan, A. F. et al. Application of the bicyclo[1.1.1]pentane motif as a nonclassical phenyl ring bioisostere in the design of a potent and orally active γ-secretase inhibitor. *J. Med. Chem.* **55**, 3414–3424 (2012).
27. Afagh, N. A. & Yudin, A. K. Chemoselectivity and the curious reactivity preferences of functional groups. *Angew. Chem. Int. Ed.* **49**, 262–310 (2010).
28. Bordwell, F. G. Equilibrium acidities in dimethyl sulfoxide solution. *Acc. Chem. Res.* **21**, 456–463 (1988).
29. Lõkov, M. et al. On the basicity of conjugated nitrogen heterocycles in different media. *Eur. J. Org. Chem.* 4475–4489 (2017).
30. Le, C. C. et al. A general small-scale reactor to enable standardization and acceleration of photocatalytic reactions. *ACS Cent. Sci.* **3**, 647–653 (2017).

Acknowledgements This research was supported by the National Institutes of Health National Institute of General Medical Sciences (R01 GM103558-03) and gifts from Merck, Bristol-Myers Squibb, Eli Lilly, Genentech and Johnson & Johnson. X.Z. is grateful for a postdoctoral fellowship from the Shanghai Institute of Organic Chemistry. We thank C. Liu and Y. Y. Loh for assistance in preparing this manuscript.

Reviewer information *Nature* thanks S. Bagley and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions Y.L. and X.Z. performed and analysed the experiments. Y.L., X.Z. and D.W.C.M. designed the experiments. Y.L., X.Z. and D.W.C.M. prepared the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0234-8.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0234-8.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to D.W.C.M.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

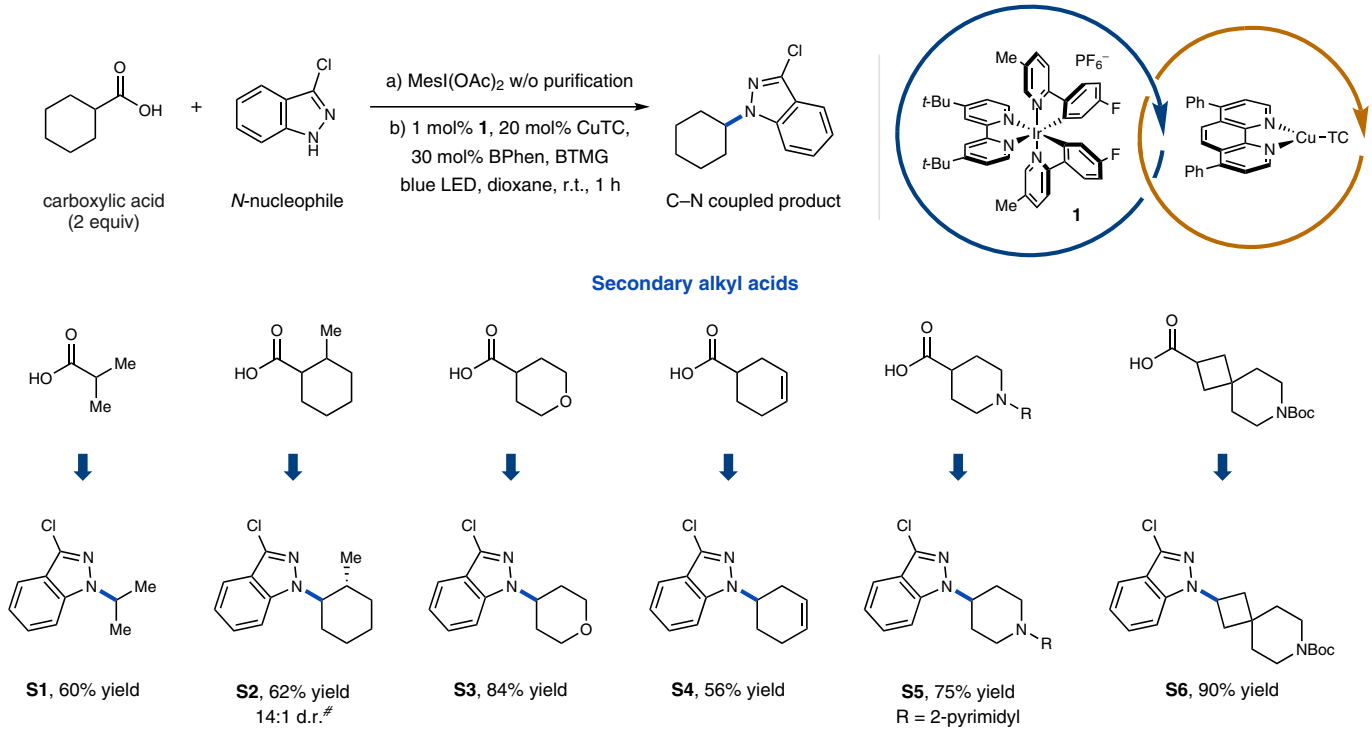
METHODS

Here we describe a typical procedure for the decarboxylative sp^3 C–N coupling reaction; a summary of general conditions is included in Supplementary Information (Supplementary Fig. 36) and further experimental details are also provided in Supplementary Information.

Procedure for decarboxylative sp^3 C–N couplings. To a 20 ml or 40 ml vial equipped with a stir bar was added photocatalyst, nitrogen nucleophile, iodomesitylene dicarboxylate, copper salt, and ligand. Dioxane was added followed by addition of the base. The solution was sonicated for 1–3 min until it became homogeneous. Next, the solution was degassed by sparging with nitrogen for 5–10 min before sealing with Parafilm. The reaction was stirred and irradiated using two 34-W blue LED lamps (3 cm away, with cooling fan to keep the reaction at room temperature) for 1 h. The reaction mixture was removed from the light, cooled to ambient temperature, diluted with water (15 ml) and ethyl acetate (25 ml), and the aqueous layer was extracted with ethyl acetate (3×25 ml). The combined organic layers were washed with brine, dried over Na_2SO_4, filtered and concentrated. The residue was purified by flash chromatography on silica gel to afford the desired decarboxylative C–N coupling product. For aniline substrates, a solution of these nitrogen nucleophiles in dioxane was used; additionally, if the iodomesitylene dicarboxylate is a liquid, its solution in dioxane was used.

Data availability. The findings of this study are available within the paper and its Supplementary Information.

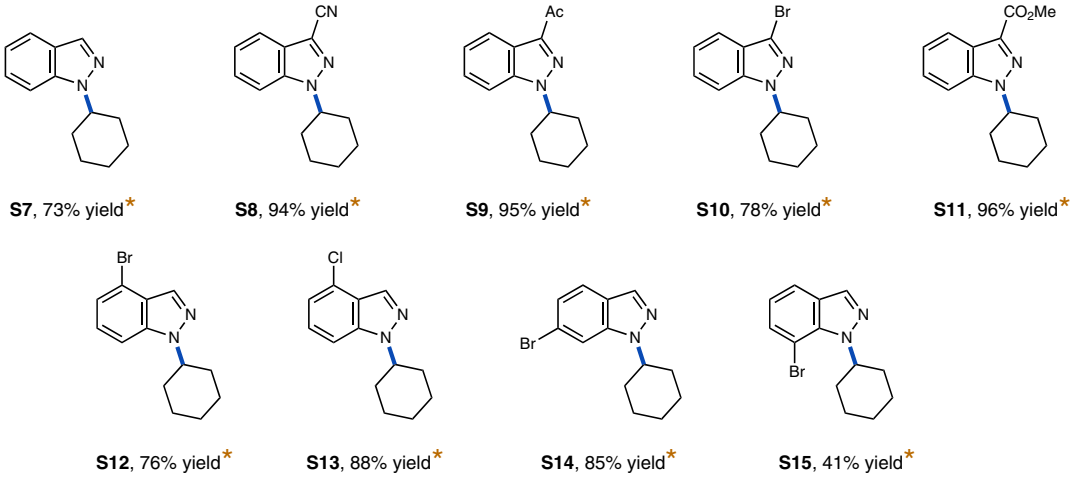
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 1 | Decarboxylative sp^3 C–N couplings with a series of secondary alkyl acids. An array of secondary alkyl carboxylic acids can be cross-coupled with 3-chloroindazole. The protocol provides the product as a single regioisomer in all cases. All yields are isolated. All reactions were replicated at least twice for consistency. See Supplementary Information for full experimental details. d.r., diastereomeric ratio. #d.r. was determined by ^{1}H NMR.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

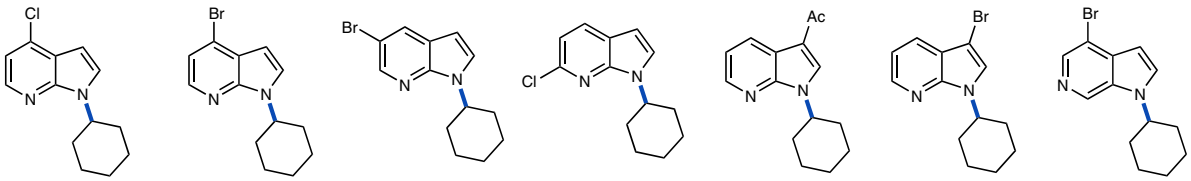
Indazoles



S7, 73% yield* **S8**, 94% yield* **S9**, 95% yield* **S10**, 78% yield* **S11**, 96% yield*

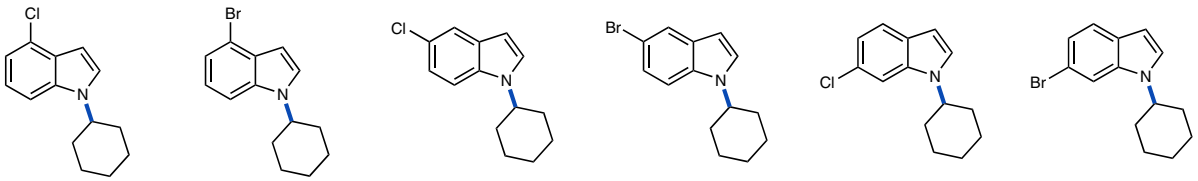
S12, 76% yield* **S13**, 88% yield* **S14**, 85% yield* **S15**, 41% yield*

Azaindoles



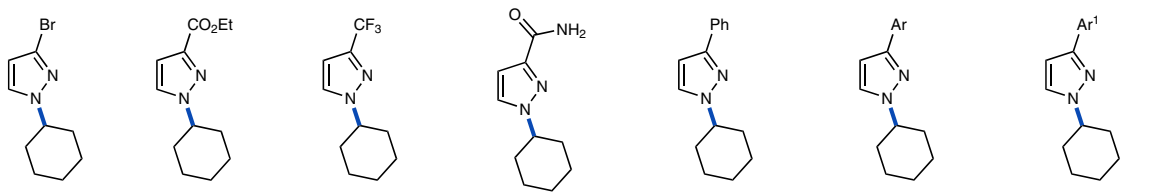
S16, 80% yield **S17**, 82% yield **S18**, 79% yield **S19**, 74% yield **S20**, 83% yield **S21**, 82% yield **S22**, 64% yield

Indoles



S23, 55% yield **S24**, 59% yield **S25**, 53% yield **S26**, 58% yield **S27**, 60% yield **S28**, 50% yield

Pyrazoles

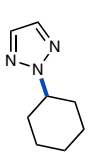


S29, 92% yield* **S30**, 72% yield* **S31**, 98% yield* **S32**, 62% yield* **S33**, 72% yield* **S34**, Ar = 4-Br-C_6H_4 85% yield* **S35**, Ar^1 = 4-Cl-C_6H_4 82% yield*

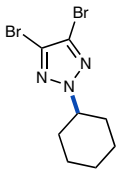
Extended Data Fig. 2 | Decarboxylative sp^3 C–N couplings with a series of nitrogen heterocycles. Various nitrogen heterocycles, including indazoles, azaindoles, indoles and pyrazoles, can cross-couple with carboxylic acids with good efficiency. All yields are isolated. All reactions were replicated at least twice for consistency. See Supplementary Information for full experimental details. *Single regioisomer.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

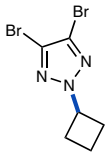
Other N-Heterocycles



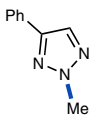
S36, 81% yield*



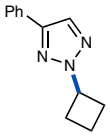
S37, 80% yield*



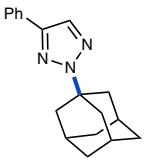
S38, 82% yield*



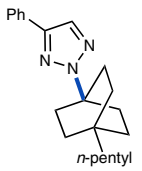
S39, 30% yield*



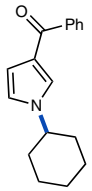
S40, 88% yield*



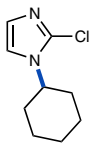
S41, 40% yield*



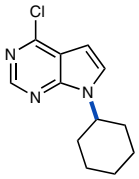
S42, 20% yield*



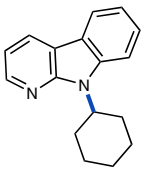
S43, 80% yield



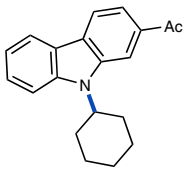
S44, 60% yield



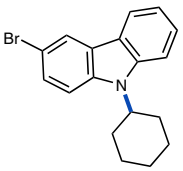
S45, 85% yield



S46, 48% yield

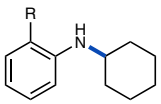


S47, 51% yield

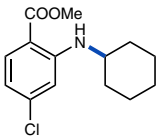


S48, 45% yield

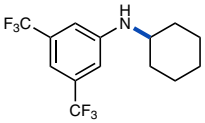
Anilines



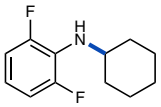
S49, R = Br, 66% yield
S50, R = Ac, 66% yield
S51, R = CN, 61% yield



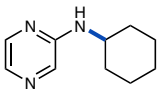
S52, 53% yield



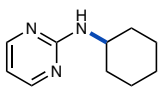
S53, 51% yield



S54, 34% yield#

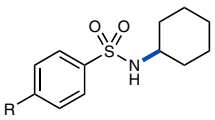


S55, 58% yield

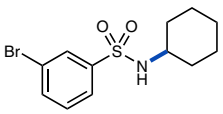


S56, 52% yield

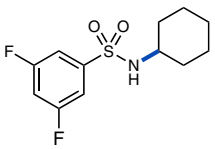
Sulfonamides



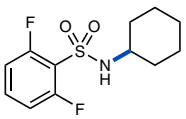
S57, R = Me, 67% yield
S58, R = F, 67% yield
S59, R = Cl, 53% yield
S60, R = Br, 63% yield



S61, 50% yield

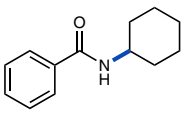


S62, 55% yield

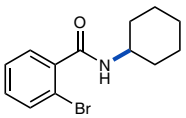


S63, 52% yield

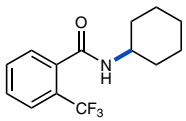
Amides



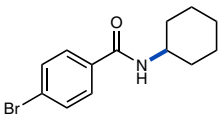
S64, 74% yield



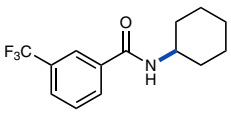
S65, 71% yield



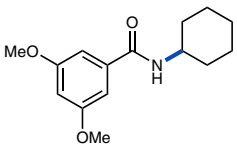
S66, 74% yield



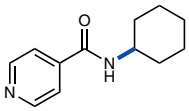
S67, 74% yield



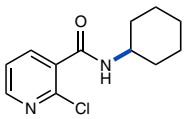
S68, 77% yield



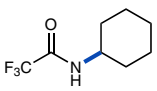
S69, 74% yield



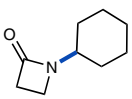
S70, 81% yield



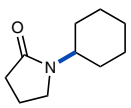
S71, 70% yield



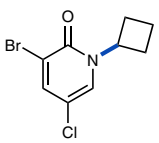
S72, 84% yield



S73, 78% yield



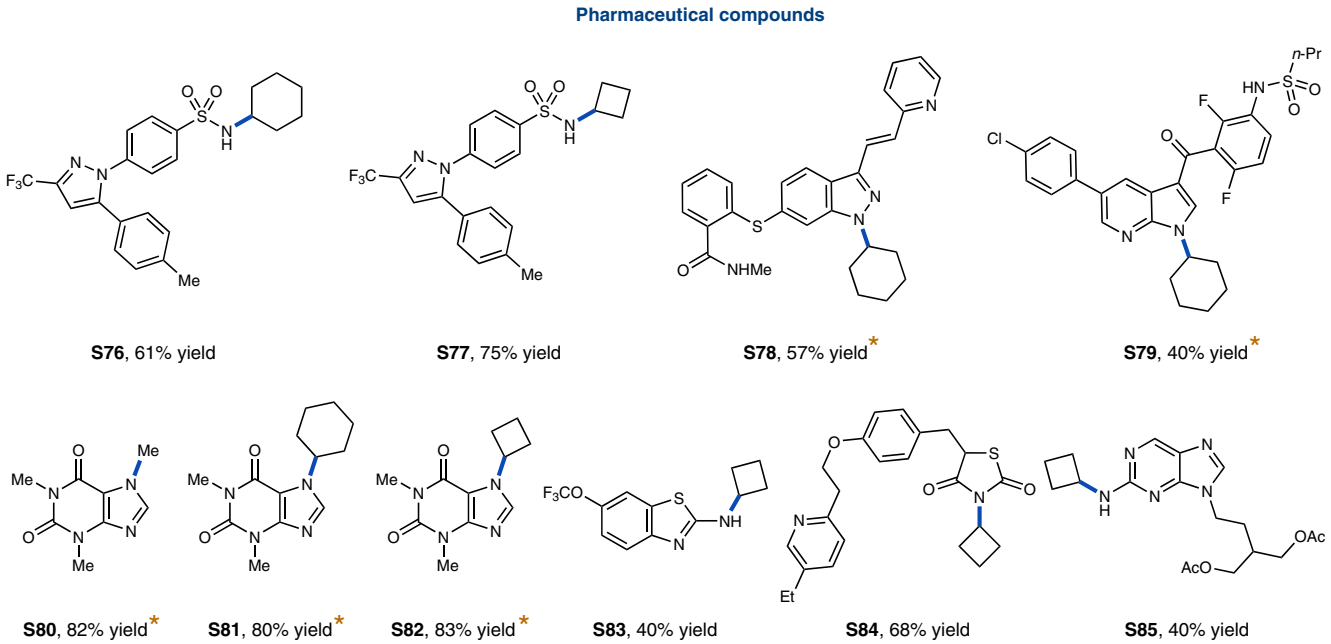
S74, 17% yield&



S75, 28% yield*

Extended Data Fig. 3 | Decarboxylative sp^3 C–N couplings with a series of nitrogen nucleophiles. Various nitrogen nucleophiles, including nitrogen heterocycles, anilines, sulfonamides and amides, can cross-couple with carboxylic acids with good efficiency. All yields are isolated unless otherwise noted. All reactions were replicated at least twice for consistency. See Supplementary Information for full experimental details. #Yield was determined by ^{19}F NMR with an internal standard. &Yield was determined by gas-chromatography analysis with an internal standard. *Single regioisomer.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 4 | Decarboxylative sp^3 C–N couplings with a series of pharmaceutical compounds. Several drug molecules can cross-couple with carboxylic acids with good efficiency. All yields are isolated. All reactions were replicated at least twice for consistency. See Supplementary Information for full experimental details. *Single regioisomer.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

https://doi.org/10.1038/s41586-018-0260-6

Spatial correlation bias in late-Cenozoic erosion histories derived from thermochronology

Taylor F. Schildgen[1,2,6]*, Pieter A. van der Beek[3,6], Hugh D. Sinclair[4] & Rasmus C. Thiede[2,5]

The potential link between erosion rates at the Earth's surface and changes in global climate has intrigued geoscientists for decades[1,2] because such a coupling has implications for the influence of silicate weathering[3,4] and organic-carbon burial[5] on climate and for the role of Quaternary glaciations in landscape evolution[1,6]. A global increase in late-Cenozoic erosion rates in response to a cooling, more variable climate has been proposed on the basis of worldwide sedimentation rates[7]. Other studies have indicated, however, that global erosion rates may have remained steady, suggesting that the reported increases in sediment-accumulation rates are due to preservation biases, depositional hiatuses and varying measurement intervals[8–10]. More recently, a global compilation of thermochronology data has been used to infer a nearly twofold increase in the erosion rate in mountainous landscapes over late-Cenozoic times[6]. It has been contended that this result is free of the biases that affect sedimentary records[11], although others have argued that it contains biases related to how thermochronological data are averaged[12] and to erosion hiatuses in glaciated landscapes[13]. Here we investigate the 30 locations with reported accelerated erosion during the late Cenozoic[6]. Our analysis shows that in 23 of these locations, the reported increases are a result of a spatial correlation bias—that is, combining data with disparate exhumation histories, thereby converting spatial erosion-rate variations into temporal increases. In four locations, the increases can be explained by changes in tectonic boundary conditions. In three cases, climatically induced accelerations are recorded, driven by localized glacial valley incision. Our findings suggest that thermochronology data currently have insufficient resolution to assess whether late-Cenozoic climate change affected erosion rates on a global scale. We suggest that a synthesis of local findings that include location-specific information may help to further investigate drivers of global erosion rates.

Thermochronology allows the reconstruction of the cooling histories of minerals, which in turn can be used to infer when and how quickly rocks were exhumed to the surface[14]. The averaging timescale of a thermochronological system is equivalent to its cooling age. That age is (to first order) the time since a mineral has cooled below its closure temperature and has begun to retain radiogenic daughter products[14]. The depth of the closure-temperature isotherm depends on both the geothermal gradient and the exhumation rate. Therefore, the different closure temperatures and resulting closure depths associated with various thermochronological systems imply that erosion rates have been averaged over different length scales and timescales[14]. To assess whether changes in erosion rates have occurred, multiple data

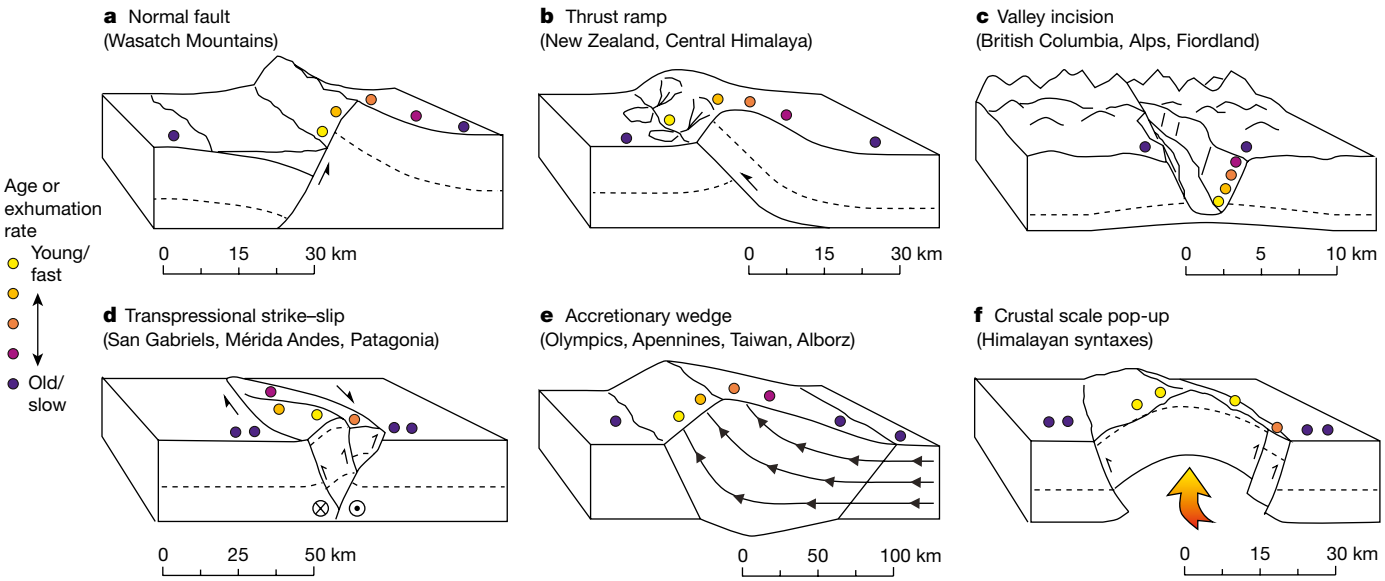


Fig. 1 | Common scenarios in which we find steep spatial gradients in thermochronological ages and exhumation rates. a–f, For each tectono-geomorphic setting, the crustal structure is sketched as a three-dimensional block and the expected thermochronological age and exhumation rate patterns are qualitatively indicated with coloured circles. Dashed lines represent the position of a crustal isochron. Half arrows show relative motion across faults. Arrows in **e** indicate particle motion pathways. The orange arrow in **f** represents upwelling middle-lower crust. The scale bar indicates the characteristic length scale of the different structures. Typical case studies discussed either in the main text or in Supplementary Information are indicated in parentheses for each tectonic setting.

[1]GFZ German Research Centre for Geosciences, Telegrafenberg, Potsdam, Germany. [2]Department of Earth and Environmental Sciences, University of Potsdam, Potsdam, Germany. [3]Université Grenoble Alpes, Institut des Sciences de la Terre (ISTerre), Grenoble, France. [4]School of Geosciences, The University of Edinburgh, Edinburgh, UK. [5]Present address: Institut für Geowissenschaften, Christian Albrecht Universität zu Kiel, Kiel, Germany. [6]These authors contributed equally: Taylor F. Schildgen, Pieter A. van der Beek. *e-mail: tschild@gfz-potsdam.de

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

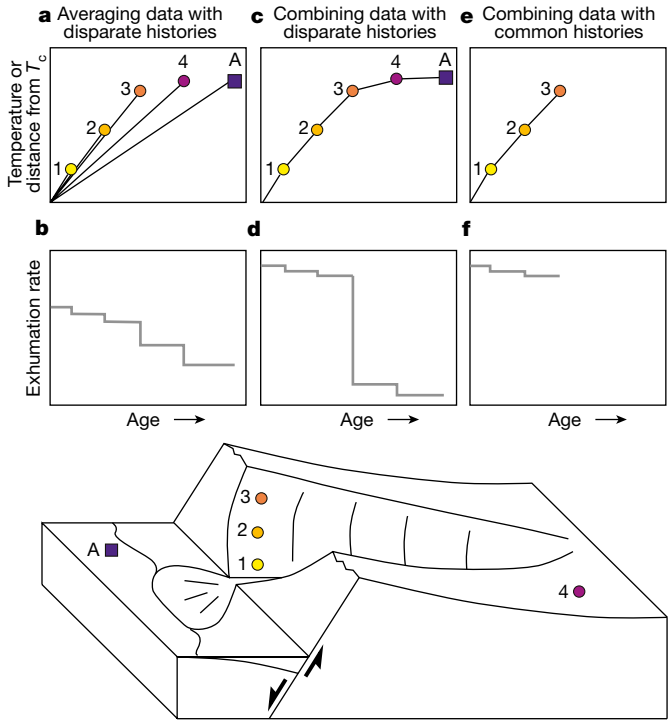


Fig. 2 | Constructing erosion histories from multiple samples or multiple thermochronometers. The lower figure illustrates thermochronological data from a tilted normal-fault-bounded block and its hanging wall. Coloured symbols correspond to thermochronological ages, as in Fig. 1. Half arrows show relative motion across the fault. **a**, **c**, **e**, Schematic age–depth (or age–temperature; see Methods) plots for the data. **b**, **d**, **f**, Erosion histories inferred by averaging data with disparate exhumation histories—a potential problem highlighted in ref. [12] (**a**, **b**); combining data with disparate exhumation histories, leading to the spatial correlation bias (and spurious increases) described here (**c**, **d**); or combining only data with common exhumation histories, resulting in correct exhumation histories (**e**, **f**). T_c, closure temperature.

points with common exhumation histories must be combined. These data points may be derived either from multiple thermochronometric systems within a single sample or from multiple samples collected from steep altitudinal transects[15]. However, if the samples are separated by some horizontal distance, they are more likely to have differing histories[16]. Several common scenarios illustrated in Fig. 1 show how thermochronological ages and exhumation rates can differ over distances of a few kilometres in regions that are tectonically deformed or deeply incised by rivers or glaciers.

The inverse relationship between erosion rates and thermochronological ages will lead to an apparent increase in erosion rates through time if erosion rates are simply averaged[12] (Fig. 2a, b). To avoid this bias in interpreting a global compilation of thermochronological data, ref. [6] used a linear inversion scheme[17] in which proximal samples were combined using a prescribed correlation length scale to derive exhumation histories (see below). However, the magnitude of the potential bias that results from combining data points that do not share a common exhumation history (for example, across a tectonic boundary), which we refer to as a 'spatial correlation bias', can be substantial. Combining rapidly exhumed samples, which tend to be younger, with slowly exhumed samples, which tend to be older, can lead to spurious accelerations in erosion rate for the analysed region (Fig. 2c, d).

To investigate the impact of spatial correlation biases on the reported acceleration in late-Cenozoic erosion[6], we first describe how a spatial correlation function, which is used to combine data points in the linear inversion model of ref. [17], is defined. Next, we conduct a review of all localities that have been reported to show an increase in late-Cenozoic erosion rates using this technique. In our review of each locality, we summarize previous interpretations and assess whether or not samples are likely to share a common exhumation history. We also evaluate the performance of the linear inversion model with a series of synthetic tests. Finally, we consider cases in which accelerated late-Cenozoic erosion appears to be unrelated to a spatial correlation bias and discuss whether climate change was the likely driver.

To construct erosion histories, the inversion model described in ref. [17] combines data points according to an assumed spatial correlation function, which gives greater weight to samples in close proximity to the location where an erosion history is determined. The spatial correlation function is based on the construction of an empirical semi-variogram, which shows the variance (mean squared difference) in erosion rates between pairs of samples, derived from calculating a steady-state erosion rate for each sample, as a function of their separation distance[18]. The empirical semi-variogram constructed for the European Alps in ref. [18], and recreated in Extended Data Fig. 1, approximates a negative exponential, with a steep increase in variance with increasing distance up to about 80 km (or an e-folding length, defined as the 'spatial correlation length', of about 25 km), after which the variance remains approximately constant. Complex spatial variations in erosion rates are expected in tectonically active areas and probably explain the high variance for the Alps data (Extended Data Fig. 1). Anisotropy in the data (that is, the data do not vary in a similar way in all directions) also contributes to the scatter. In datasets that contain anisotropy, different correlation functions are typically applied in different directions.

In the global analysis performed in ref. [6], the spatial correlation function used to combine and invert the data is an isotropic negative exponential with a uniform spatial correlation length of 30 km. In this model, data points close to a grid node are more strongly weighted when constructing an erosion history for the node's location; however, data points up to about 90 km away may still be used to derive an erosion history if no closer data points are available. Additional empirical semi-variograms were not reported to evaluate whether or not a negative exponential characterizes all parts of the global dataset, nor were the data evaluated for anisotropy.

A small subset (32 regions) of the global data compiled in ref. [6] produced erosion rates that were reported to be well resolved in the time bins from 6 to 4 million years (Myr) ago and from 2 to 0 Myr ago. The authors used these 32 regions (Extended Data Fig. 2) to assess global patterns in late-Cenozoic erosion-rate changes. Out of those regions, 27 produced erosion rates that are well resolved across each of four 2-Myr time bins from 8 to 0 Myr ago, which were in turn used to infer a nearly twofold late-Cenozoic increase in erosion rates across mountainous landscapes worldwide[6]. For each of the 32 regions with erosion histories that were reported to be well resolved (30 of which show accelerations), we conduct a literature review to compare the original data and interpretations to the erosion-rate histories reported in ref. [6]. The regions include areas affected by normal faulting, transpression, thrust ramps, crustal pop-up structures and accretionary orogenic wedges. Other regions were affected by headward migration of river incision and glacial erosion. In each case, we explore whether the assumed spatial correlation function is justified, specifically considering whether the data points have a common exhumation history. Two of the regions that we reviewed are presented below; these were chosen because they represent an important part (21%) of the dataset that was used to infer a global increase in late-Cenozoic erosion rates[6] and they are examples of common sources of the spatial correlation bias. Further discussion and reviews of all remaining regions can be found in Supplementary Information.

Our first case study comprises the western European Alps. The thermochronological database for this region defines two tectonic domains: the external zone, in which thermochronological ages are relatively young (<10 Myr to 15 Myr for apatite fission-track (AFT) ages; <20 Myr for zircon fission-track (ZFT) ages), and the internal zone, where AFT and ZFT ages are typically >20 Myr (Fig. 3a). The limit between these two zones corresponds to the Penninic frontal thrust (PFT), which coincides with a steep spatial gradient in thermochronological ages[19–21]. Interpretations of the data from the internal zone

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

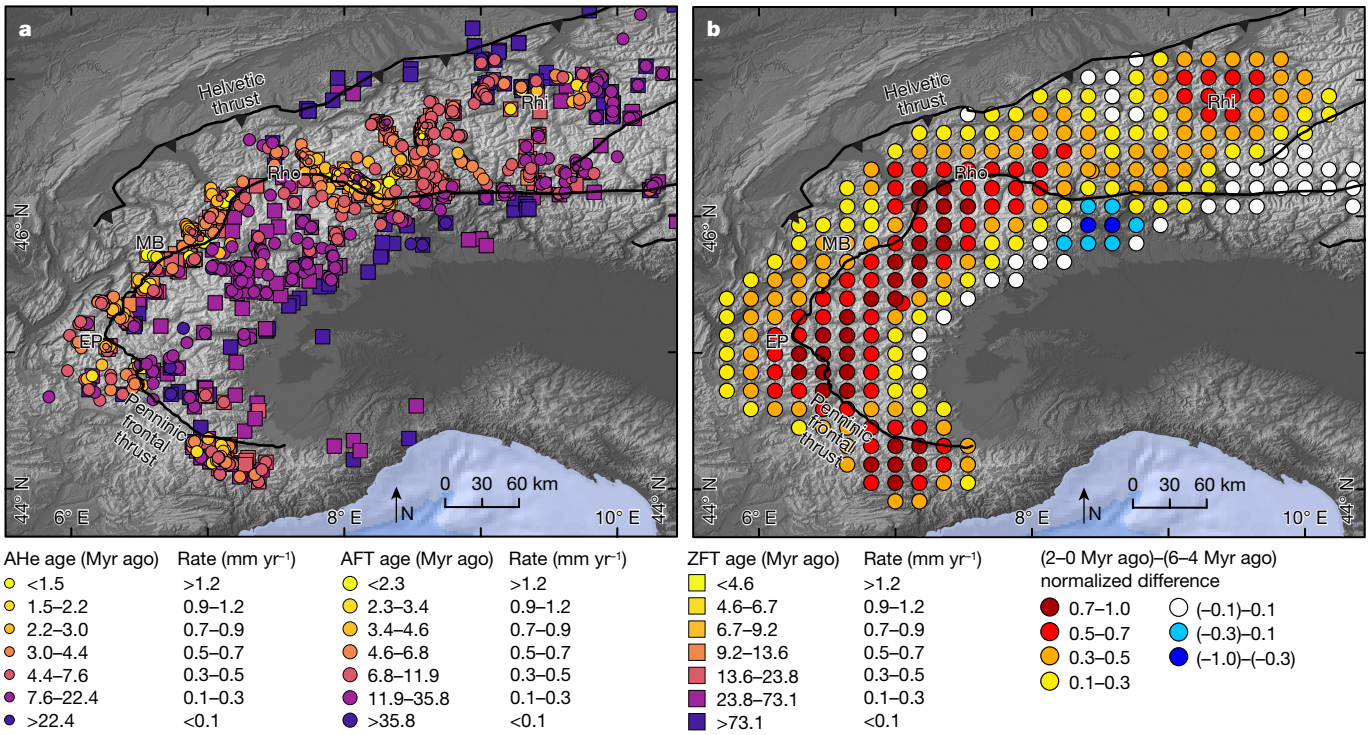


Fig. 3 | Thermochronology data and modelled erosion-rate changes for the western European Alps. a, Thermochronology data are derived from 52 different sources compiled by refs [6,25]. Equivalent colours for thermochronometers correspond to equivalent one-dimensional steady-state erosion rates (see Methods for details). **b**, Normalized difference (see Methods) of erosion rates resolved from 2–0 Myr ago compared to those from 6–4 Myr ago, from the supplementary information of ref. [6]. AHe, apatite (U–Th)/He system; MB, Mont Blanc massif; EP, Ecrins–Pelvoux massif; Rho, Rhône Valley; Rhi, Rhine Valley.

invoke steady, slow exhumation since early Miocene times, after a phase of rapid exhumation in the Oligocene[21,22]. By contrast, the external crystalline massifs in the external zone show rapid exhumation in late Miocene–early Pliocene times, followed by slower rates and finally localized exhumation of valley bottoms; the latter is attributed to glacial valley incision since the mid-Pleistocene[23].

In contrast to these interpretations, linear inversion models[6,18,24] show a broad band of erosion-rate increases in the western Alps during the Quaternary, which were attributed to a combination of slab breakoff and glacial erosion (Fig. 3b). An analogous earlier analysis reached similar conclusions[25]. However, the band of increased erosion rates straddles the PFT (Fig. 3b), which marks the boundary between two blocks that have experienced contrasting exhumation histories. The inferred erosion-rate increases occur throughout the internal zone, east of the PFT, where no thermochronological ages younger than 4 Myr occur. Thus, the data cannot resolve any variations in erosion rates later than 4 Myr ago. Our synthetic tests show that the reported increases in erosion rates result from a combination of data across the PFT (Extended Data Fig. 3). When we impose different exhumation rates across this tectonic boundary but hold them constant through time, spurious increases are produced by the linear inversion in both the external and the internal Alps (see Methods, Extended Data Fig. 3).

In our second case study, the Mount Cook region of the Southern Alps in New Zealand (Fig. 4, Extended Data Fig. 4), oblique convergence between the Australian and Pacific plates is concentrated along the Alpine fault. This region has long been considered as a type example of a doubly vergent thrust wedge in which exhumation is concentrated in the retro-wedge owing to both the tectonic and the erosional asymmetry of the system[26,27]. Rocks are exhumed from progressively greater depths towards the Alpine fault. As a result, metamorphic grade increases and thermochronological ages decrease systematically towards the fault[28,29]. Each thermochronological system shows young (reset) ages close to the Alpine fault; the width of the reset age zone depends on the dip of the fault and the closure temperature of the system (Fig. 4a–d). Thermo-mechanical[27] and thermo-kinematic[29,30] modelling has shown that the age pattern can be explained by steady convergence at rates of around 10 mm yr^{-1} since about 5–6 Myr ago[28], consistent with plate-tectonic reconstructions indicating a switch from pure strike–slip to oblique convergence along the Alpine fault at that time[31].

In contrast to these earlier findings, the analysis reported in ref. [6] suggests a 20-fold increase in late-Cenozoic erosion rates along the crest of the Southern Alps. The increases are only resolved in a zone between about 20 km and 30 km southeast of the Alpine fault. Closer to the fault, all ages are considerably younger than 6 Myr, leading to loss of resolution for the older age bins, whereas farther away from the fault, all ages are substantially older than 6 Myr, leading to loss of resolution in the younger age bins. The inferred increases in erosion rates are due to the systematic spatial variations in ages, which reflect the laterally varying exhumation pathways, but are translated by the linear inversion model into temporal increases in erosion rates, as illustrated in another synthetic test (see Methods, Extended Data Fig. 5).

Quantifying how much of the proposed worldwide increase in late-Cenozoic mountain erosion rates results from a true increase recorded by samples that share a common exhumation history, rather than from a bias linked to the combination of samples with disparate exhumation histories, is a challenge. We have attempted to do so qualitatively by summarizing our conclusions from all of the areas for which increases in late-Cenozoic erosion rates were inferred[6]. We find that the reported increases can be explained by the combination of data with disparate exhumation histories in 23 cases. Increases in four cases can be explained by changes in tectonic boundary conditions, two are related to both changes in tectonics and glacial activity, and one appears to be associated with glacial valley incision alone (Extended Data Table 1, Extended Data Fig. 2). Hence, only in three regions do thermochronology data reflect an impact of late-Cenozoic cooling.

Theoretically, regional changes in exhumation rates could be correctly resolved using linear modelling approaches in regions that have

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

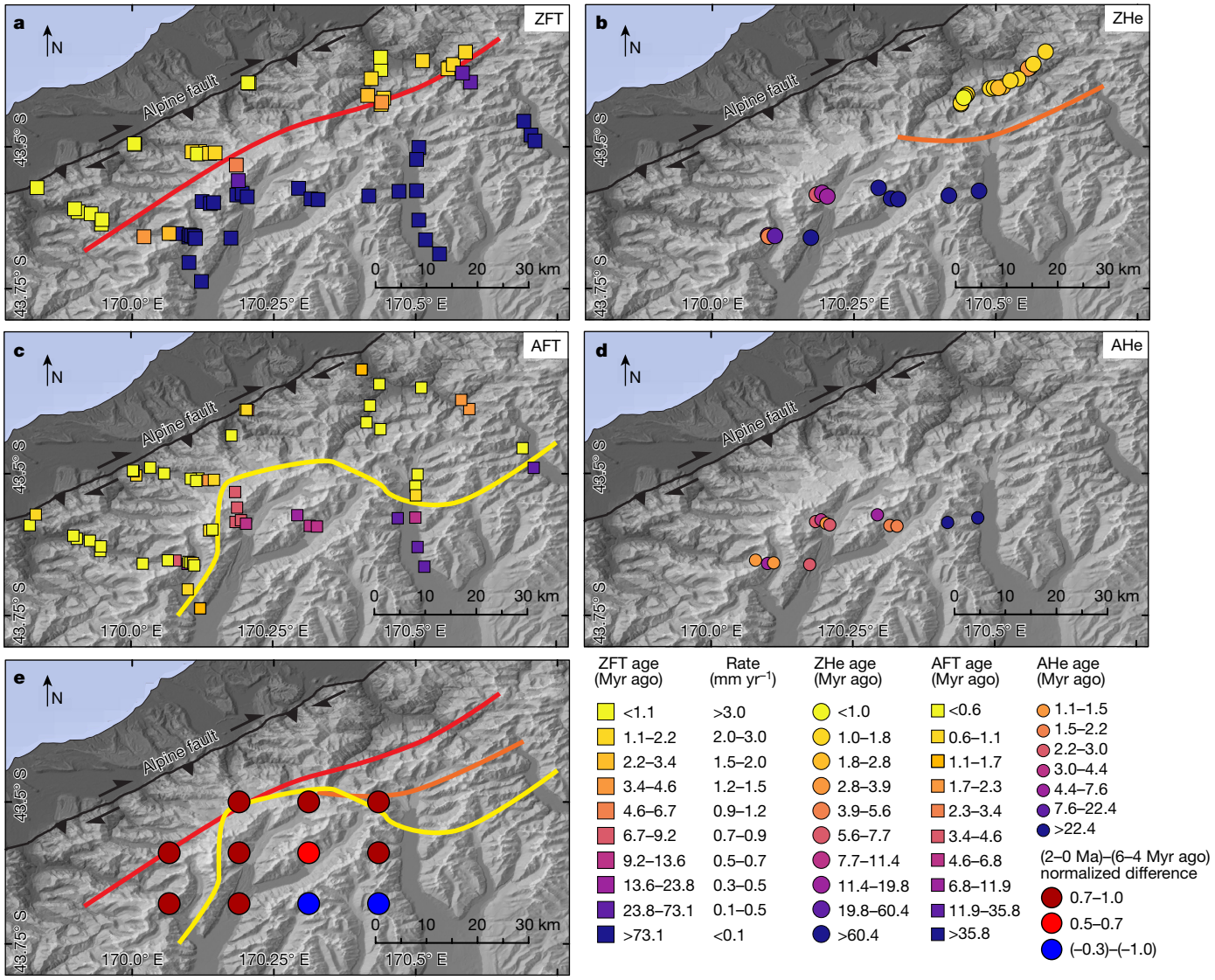


Fig. 4 | Thermochronology data and modelled erosion-rate changes for the Mount Cook region in New Zealand. a–d, Thermochronology data from refs [28,33] (AFT, ZFT), ref. [29] (ZHe) and ref. [30] (AHe, AFT, ZHe, ZFT). Equivalent colours for thermochronometers correspond to equivalent one-dimensional steady-state erosion rates. Coloured lines in **a**–**c** and **e** mark the approximate boundaries between reset and un-reset thermochronometer ages (red, ZFT; orange, ZHe; yellow, AFT). AHe, apatite (U–Th)/He system; ZHe, zircon (U–Th)/He system. **e**, Normalized difference of erosion rates resolved from 2–0 Myr ago compared to the those from 6–4 Myr ago, from the supplementary information of ref. [6].

not experienced changes in short-wavelength relief, with data available from either (1) multiple thermochronometers analysed from the same sample in an area of mostly vertical exhumation or (2) steep elevation transects of samples that do not cross spatial gradients in exhumation[17]. Reducing the spatial correlation length in the linear inversion model results in fewer cases of unwarranted data combination, as illustrated in our synthetic tests (see Methods and Extended Data Figs. 3, 5, 6). However, as the correlation length is decreased, fewer areas yield well-resolved erosion histories. More problematically, our synthetic tests show that the spatial correlation bias can create spurious increases even in areas of high data density, multiple thermochronometers available from individual samples and steep sample transects (Extended Data Fig. 3).

On the basis of our literature review and synthetic tests, we find that thermochronology data do not resolve a worldwide late-Cenozoic increase in mountain erosion rates. Both the magnitude of the inferred increase and its purported occurrence throughout mountainous landscapes appear to be largely artefacts related to combining data with disparate exhumation histories. These results, combined with previous analyses of bias in the sedimentary record[8–10], call into question the evidence presented to date for a worldwide late-Cenozoic increase in erosion rates.

Considering the pitfalls of a global approach to data analysis that neglects the local context of the data, an approach that synthesizes detailed local work and includes location-specific aspects in data analysis is probably our best way forward. Such an approach, as summarized here, shows only a few locations in which resolved increases in late-Cenozoic erosion rates are unambiguously linked to late-Cenozoic climate change, and in those areas, such increases are mostly restricted to localized glacial valley incision[23,32]. Hence, even in cases where data do resolve late-Cenozoic changes in landscape relief, the importance of such changes with respect to global sediment and carbon budgets must be carefully considered.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0260-6.

Received: 15 October 2017; Accepted: 8 May 2018;
Published online 4 July 2018.

1. Molnar, P. & England, P. Late Cenozoic uplift of mountain ranges and global climate change: chicken or egg? *Nature* **346**, 29–34 (1990).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

2. Raymo, M. E. & Ruddiman, W. F. Tectonic forcing of late Cenozoic climate. *Nature* **359**, 117–122 (1992).
3. Berner, R. A., Lasaga, A. C. & Garrels, R. M. The carbonate-silicate geochemical cycle and its effect on atmospheric carbon dioxide over the past 100 million years. *Am. J. Sci.* **283**, 641–683 (1983).
4. Kump, L. R., Brantley, S. L. & Arthur, M. A. Chemical weathering, atmospheric CO_2, and climate. *Annu. Rev. Earth Planet. Sci.* **28**, 611–667 (2000).
5. France-Lanord, C. & Derry, L. A. Organic carbon burial forcing of the carbon cycle from Himalayan erosion. *Nature* **390**, 65–67 (1997).
6. Herman, F. et al. Worldwide acceleration of mountain erosion under a cooling climate. *Nature* **504**, 423–426 (2013).
7. Zhang, P., Molnar, P. & Downs, W. R. Increased sedimentation rates and grain sizes 2–4 Myr ago due to the influence of climate change on erosion rates. *Nature* **410**, 891–897 (2001).
8. Sadler, P. M. The influence of hiatuses on sediment accumulation rates. *Geores. Forum* **5**, 15–40 (1999).
9. Schumer, R. & Jerolmack, D. J. Real and apparent changes in sediment deposition rates through time. *J. Geophys. Res.* **114**, F00A06 (2009).
10. Willenbring, J. K. & von Blanckenburg, F. Long-term stability of global erosion rates and weathering during late-Cenozoic cooling. *Nature* **465**, 211–214 (2010).
11. Herman, F. & Champagnac, J.-D. Plio-Pleistocene increase of erosion rates in mountain belts in response to climate change. *Terra Nova* **28**, 2–10 (2016).
12. Willenbring, J. K. & Jerolmack, D. J. The null hypothesis: globally steady rates of erosion, weathering fluxes and shelf sediment accumulation during Late Cenozoic mountain uplift and glaciation. *Terra Nova* **28**, 11–18 (2016).
13. Ganti, V. et al. Time scale bias in erosion rates of glaciated landscapes. *Sci. Adv.* **2**, e1600204 (2016).
14. Reiners, P. W. & Brandon, M. T. Using thermochronology to understand orogenic erosion. *Annu. Rev. Earth Planet. Sci.* **34**, 419–466 (2006).
15. Hurford, A. J. Uplift and cooling pathways derived from fission track analysis and mica dating: a review. *Geol. Rundsch.* **80**, 349–368 (1991).
16. Stephenson, J., Gallagher, K. & Holmes, C. C. Low temperature thermochronology and strategies for multiple samples 2: partition modelling for 2D/3D distributions with discontinuities. *Earth Planet. Sci. Lett.* **241**, 557–570 (2006).
17. Fox, M., Herman, F., Willett, S. D. & May, D. A. A linear inversion method to infer exhumation rates in space and time from thermochronometric data. *Earth Surf. Dynam.* **2**, 47–65 (2014).
18. Fox, M., Herman, F., Willett, S. D. & Schmid, S. M. The exhumation history of the European Alps inferred from linear inversion of thermochronometric data. *Am. J. Sci.* **316**, 505–541 (2016).
19. Seward, D. & Mancktelow, N. S. Neogene kinematics of the central and western Alps: evidence from fission-track dating. *Geology* **22**, 803–806 (1994).
20. Fügenschuh, B. & Schmid, S. M. Late stages of deformation and exhumation of an orogen constrained by fission-track data: a case study in the Western Alps. *Geol. Soc. Am. Bull.* **115**, 1425–1440 (2003).
21. Tricart, P., van der Beek, P., Schwartz, S. & Labrin, E. Diachronous late-stage exhumation across the western Alpine arc: constraints from apatite fission-track thermochronology between the Pelvoux and Dora-Maira Massifs. *J. Geol. Soc. Lond.* **164**, 163–174 (2007).
22. Beucher, R., van der Beek, P., Braun, J. & Batt, G. E. Exhumation and relief development in the Pelvoux and Dora-Maira massifs (western Alps) assessed by spectral analysis and inversion of thermochronological age transects. *J. Geophys. Res.* **117**, F03030 (2012).
23. Valla, P. G., Shuster, D. L. & van der Beek, P. A. Significant increase in relief of the European Alps during mid-Pleistocene glaciations. *Nat. Geosci.* **4**, 688–692 (2011).
24. Fox, M., Herman, F., Kissling, E. & Willett, S. D. Rapid exhumation in the Western Alps driven by slab detachment and glacial erosion. *Geology* **43**, 379–382 (2015).
25. Vernon, A. J., van der Beek, P. A., Sinclair, H. D. & Rahn, M. K. Increase in late Neogene denudation of the European Alps confirmed by analysis of a fission-track thermochronology database. *Earth Planet. Sci. Lett.* **270**, 316–329 (2008).
26. Beaumont, C., Kamp, P. J. J., Hamilton, J. & Fullsack, P. The continental collision zone, South Island, New Zealand: comparison of geodynamical models and observations. *J. Geophys. Res.* **101**, 3333–3359 (1996).
27. Batt, G. E. & Braun, J. The tectonic evolution of the Southern Alps, New Zealand: insights from fully thermally coupled dynamical modelling. *Geophys. J. Int.* **136**, 403–420 (1999).
28. Batt, G. E., Braun, J., Kohn, B. P. & McDougall, I. Thermochronological analysis of the dynamics of the Southern Alps, New Zealand. *Geol. Soc. Am. Bull.* **112**, 250–266 (2000).
29. Herman, F., Braun, J. & Dunlap, W. J. Tectonomorphic scenarios in the Southern Alps of New Zealand. *J. Geophys. Res.* **112**, B04201 (2007).
30. Herman, F., Cox, S. C. & Kamp, P. J. J. Low-temperature thermochronology and thermokinematic modeling of deformation, exhumation, and development of topography in the central Southern Alps, New Zealand. *Tectonics* **28**, TC5011 (2009).
31. Walcott, R. I. Modes of oblique compression: Late Cenozoic tectonics of the South Island of New Zealand. *Rev. Geophys.* **36**, 1–26 (1998).
32. Ehlers, T. A., Farley, K. A., Rusmore, M. E. & Woodsworth, G. J. Apatite (U-Th)/He signal of large-magnitude accelerated glacial erosion, southwest British Columbia. *Geology* **34**, 765–768 (2006).
33. Tippett, J. M. & Kamp, P. J. J. Fission track analysis of the Late Cenozoic vertical kinematics of continental Pacific crust, South Island, New Zealand. *J. Geophys. Res.* **98**, 16119–16148 (1993).

Acknowledgements We thank P. Valla for discussions on the topic. D. Jerolmack and A. Ault provided feedback on an early version of the manuscript. T.F.S. acknowledges support from the Emmy Noether Program of the Deutsche Forschungsgemeinschaft (DFG), grant number SCHI 1241/1-1, and the Helmholtz Association's Initiative and Networking Fund. P.A.v.d.B acknowledges support from the Institut Universitaire de France (IUF). ISTerre is part of Labex OSUG@2020 (ANR10 LABX56). R.C.T. was funded by the DFG grant number TH 1371/5–1.

Reviewer information *Nature* thanks R. Flowers, E. Kirby and J. Spotila for their contribution to the peer review of this work.

Author contributions T.F.S., P.A.v.d.B. and H.D.S. discussed early versions of the ideas presented here, which were refined following discussion among all the authors. T.F.S., P.A.v.d.B. and R.C.T. compiled data for the figures. T.F.S. and P.A.v.d.B. created all figures and constructed the *edot2age.m* Matlab script. P.A.v.d.B. carried out the synthetic model runs with the help of T.F.S. T.F.S., P.A.v.d.B. and H.D.S. analysed individual case studies. All authors discussed the results and contributed to the writing of the text.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0260-6.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0260-6.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to T.F.S.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Obtaining normalized-difference maps of erosion rates from the global analysis. The full results of ref. [6] are provided in a 207-page data table in the supplementary information of that paper. That table provides the longitude, latitude, erosion rate and resolution for each of the grid points in the model[6], divided into time bins (that is, in 2-Myr increments between 8 Myr ago and the present). The analysis of ref. [6] employed ratios of erosion rates from 2–0 Myr ago and compared to 6–4 Myr ago to assess global patterns of erosion-rate changes. A disadvantage of ratios (final rate divided by initial rate) is that the values increase substantially for very slow initial erosion rates. To avoid this problem, we instead calculate normalized differences (ND) between the final erosion rates (e_f, from 2–0 Myr ago) and initial erosion rates (e_i, from 6–4 Myr ago):

$$\mathrm{ND} = \frac{e_f - e_i}{\max(e_f, e_i)} \quad (1)$$

The difference between the final and initial erosion rate is divided by the maximum of the two rates. Thus, increases are scaled between 0 and 1 and decreases are scaled between 0 and −1. These values track fractional changes in erosion rates, so an increase in erosion rates from 0.5 mm yr^{-1} to 1 mm yr^{-1} yields the same normalized difference as an increase from 0.05 mm yr^{-1} to 0.1 mm yr^{-1}. Increases and decreases are also symmetric, so an increase from 0.05 mm yr^{-1} to 0.1 mm yr^{-1} has a normalized difference of 0.5, whereas a decrease from 0.1 mm yr^{-1} to 0.05 mm yr^{-1} has a normalized difference of −0.5. Calculating these values for each 'resolved' location (following the definition in ref. [6]) was automated using a Matlab script. The data were then exported to ArcMap GIS (Esri) to be plotted in map view. Such maps were not shown in ref. [6], but they are critical for assessing the quality of the model results; all the data needed to make those maps can be found in the supplementary information of ref. [6].

Finding the steady-state thermochronological age corresponding to a steady erosion rate. When constructing maps for each case study, we coloured each thermochronological data point according to its corresponding one-dimensional steady-state erosion rate. We emphasize that these rates are calculated only to facilitate visual comparison of data from different thermochronometers; the rates do not reflect an attempt to reconstruct detailed exhumation histories. If the data points of multiple thermochronometers from a small region have the same colour, the data are consistent with minimal changes in exhumation rates through time. Hence, this approach provides a simple way to visually assess mean, time-averaged exhumation rates from multiple thermochronological systems. However, because our predicted erosion rates do not consider sample elevation, any comparison of samples should be among those at a similar elevation.

In detail, we determined what age corresponds to a given exhumation rate for each thermochronometer (Extended Data Table 2) and coloured data in each map according to a range of specified rates. Changing the parameters in this model will modify the absolute rates, but not the relative rates, which are of interest here. In the following, we describe the steps used to make the conversion from erosion rate to age on the basis of the methodology developed in refs [34,35], which are encoded in the Matlab script *edot2age.m* (available in Supplementary Information).

The age of a thermochronological system is, to the first order, equal to the time since that system crossed its closure-temperature isotherm as it was exhumed to the surface. The closure temperature (T_c) of a monotonously cooling thermochronological system can be modelled as[36]:

$$T_c = \frac{E_a}{R\ln(A\tau D_0/a^2)} \quad (2)$$

where R is the universal gas constant, A is a geometry factor, E_a, D_0 and a are experimentally determined diffusion parameters and τ is a characteristic time that depends on the cooling rate, $\delta T/\delta t$, where t is the time:

$$\tau = -\frac{RT_c^2}{E_a(\delta T/\delta t)} \quad (3)$$

We calculate the closure temperatures for the apatite and zircon (U–Th)/He systems (AHe and ZHe, respectively) and for the AFT and ZFT systems as a function of cooling rate using diffusion parameters given in ref. [14] (see Extended Data Table 3). We note that although AFT and ZFT annealing cannot strictly be treated as linear Arrhenius diffusion[37], the above equations provide a good first-order approximation for this process as well, as discussed in ref. [14].

To transform the closure temperature into a closure depth, the geothermal gradient needs to be known. The equation describing the temperature, T, (in kelvin) as a function of depth, z, (in kilometres) in a one-dimensional system including vertical rock advection (but no heat production) is[35,38]:

$$T(z) = T_0 + (T_L - T_0)\frac{1-e^{-z\varepsilon/\kappa}}{1-e^{-L\varepsilon/\kappa}} \quad (4)$$

where L (km) is the 'characteristic length' of the system (in our case, the crustal thickness), ε (km yr^{-1}) is the vertical exhumation rate, κ (km^2 yr^{-1}) is the thermal diffusivity, and T_0 and T_L (K) are the temperatures at the top and the base of the system (at depths 0 and L), respectively.

We can rearrange this equation to give the depth as a function of temperature (that is, to derive a closure depth from a closure temperature):

$$z(T) = \frac{\kappa}{\varepsilon}\ln\left[1 - \frac{T-T_0}{T_L-T_0}(1-e^{-L\varepsilon/\kappa})\right] \quad (5)$$

We take the derivative of equation (4) with respect to depth to obtain the geothermal gradient:

$$\frac{dT}{dz} = \frac{\varepsilon(T_L-T_0)}{\kappa(1-e^{-L\varepsilon/\kappa})}e^{-z\varepsilon/\kappa} \quad (6)$$

Equations (2)–(6) are solved iteratively. The geothermal gradient is used together with the input exhumation rate to calculate the cooling rate at the closure depth and thereby the closure temperature. The iteration quickly converges to a steady-state geothermal gradient, closure temperature and depth, and thermochronologic age as a function of exhumation rate.

Synthetic tests of the linear inversion model. We conducted synthetic tests to evaluate whether the spatial correlation bias may exist in various tectonic settings and to test the impact of different values for the spatial correlation length and the a priori erosion rate used in the model inversion. To create synthetic datasets, we extracted elevation data and sample locations from three different areas: the western European Alps, the Mount Cook area of New Zealand and the Wasatch Mountains. We predicted the ages at the sample locations by running forward models with the three-dimensional thermal-kinematic code Pecube[39], assuming spatially variable but temporally constant exhumation rates. In these forward models, we used exhumation rates that produce a range of ages similar to what is found in the real data in each area. We added a random 10% error to these ages to obtain a more natural spread and then applied the linear inversion model of ref. [17] to the predicted ages. For both the thermal–kinematic forward model and the linear inversion, we used the same thermal and thermochronometric parameters as in ref. [6]; these are summarized in Extended Data Table 3. For each synthetic test, we show maps (Extended Data Figs. 3, 5, 6) of the input erosion rates (constant in time) and the erosion rates predicted by the inversion model from 6–4 Myr ago and from 2–0 Myr ago. Model-predicted maps include contour lines showing the resolution. We also present maps showing the normalized difference between the erosion rates of the two time windows. Importantly, because the input exhumation rates are constant in time for all our model runs, any predicted change between the erosion rates of different time bins is an artefact of the linear inversion model.

For the synthetic test of the western European Alps, we extracted the topography and sample locations from a portion of the range spanning the northeast–southwest striking, sub-linear segment of the Penninic frontal thrust (PFT) between the Ecrins–Pelvoux massif and the Rhône valley (Fig. 3). We modelled the PFT as a vertical tectonic boundary separating two regions of contrasting but temporally constant exhumation rates, and we set the exhumation rates for the external and internal zones to be 1.0 mm yr^{-1} and 0.25 mm yr^{-1}, respectively (Extended Data Fig. 3a). We ran the model at these rates for 20 Myr before predicting ages for various thermochronometers (AHe, AFT and ZFT) at the sample locations for which real data are available. Our choice to use real data locations reflects our aim to perform the linear inversion with a realistic distribution of data points. Predicted AHe ages (18 in total) range from 2.0 Myr to 3.3 Myr in the external zone; a single age in the internal zone is 8.6 Myr. Predicted AFT ages (176 in total) range from 2.5 Myr to 5.6 Myr in the external zone and from 11.3 Myr to 22.5 Myr in the internal zone. Predicted ZFT ages (114 in total) range from 6.4 Myr to 8.8 Myr in the external zone; in the internal zone, the ZFT ages are unreset (that is, they have not been exhumed from deep enough to yield ages representative of the imposed exhumation rate) and they range from 32.5 Myr to 34.5 Myr (Extended Data Fig. 3b). Using a spatial correlation length of 30 km (as used in ref. [6]), the linear inversion of the data predicts 'resolved' increases in exhumation rates through time across an approximately 110-km-wide zone (using the resolution cut-off of 0.25, as in ref. [6]) that spans both sides of the PFT; that is, both the internal and the external zones (Extended Data Fig. 3e). Using a correlation length of only 10 km, the inversion results in a narrower (about 30-km-wide) zone of resolved increases, but those increases still span both the internal and external zones (Extended Data Fig. 3h). The decreased area of resolved changes for the 10 km correlation length occurs because fewer samples are combined when determining exhumation histories. This synthetic test illustrates the strength of the spatial correlation bias, even in regions

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

with a high data density, multiple thermochronometers available for individual samples, and steep age-elevation transects.

For the New Zealand synthetic test, we extracted topographic data and sample locations from the Mount Cook area (Fig. 4). We predicted exhumation rates from a model that assumes overthrusting for 10 Myr at a constant rate of 5 mm yr^{-1} along a fault with a geometry taken from ref. [30]: dipping at 45° from the surface to a depth of 15 km and at 9.5° below that depth, before soling out in a horizontal detachment at a depth of 20 km. Predicted exhumation rates associated with this setup vary stepwise, from 2.2 mm yr^{-1} adjacent to the fault to zero about 47 km away from it (Extended Data Fig. 5a). From this model, we predict 13 AHe ages between 1.0 Myr and 5.4 Myr, 50 reset AFT ages between 0.8 Myr and 5.4 Myr, 19 reset and 3 partially reset ZHe ages between 1.7 Myr and 7.7 Myr and between 10 Myr and 34 Myr, respectively, and 30 reset ZFT ages between 1.8 Myr and 5.8 Myr. In total, 7 AFT, 5 ZHe and 33 ZFT ages are predicted to be unreset and vary from 80 Myr to 100 Myr (Extended Data Fig. 5b). The linear inversion of the data using a 30 km spatial correlation length predicts very low rates (close to the a priori erosion rate of 0.35 mm yr^{-1}) from 6–4 Myr ago throughout the model domain, with a predicted resolution greater than 0.25 in a large part of it (Extended Data Fig. 5c). From 2–0 Myr ago, the model predicts rates that increase towards the fault (Extended Data Fig. 5d); as a result, large and resolved increases (normalized differences ranging between 0.7 and 0.9) are predicted over most of the model domain (Extended Data Fig. 5e). Choosing a shorter correlation length of 10 km leads to similar results, but with lower resolution, in particular from 6–4 Myr ago, leading to predicted resolved increases over a smaller part of the model domain (Extended Data Fig. 5h).

For the Wasatch synthetic test, we extracted the topography and sample locations from the portion of the mountain range for which thermochronological data are available (Supplementary Fig. 1a). We set the exhumation rate on the left side of the model (along the Wasatch fault) to 1.2 mm yr^{-1} and on the right side of the model to zero, creating a smooth spatial variation in exhumation rates across the model domain. We ran Pecube at these rates for 12 Myr before predicting AHe and AFT ages at the sample locations where real data are available (Extended Data Fig. 6). Twenty-seven predicted AHe ages range from 1.9 Myr to 7.3 Myr, 33 reset AFT ages range from 2.7 Myr to 10 Myr, whereas 4 unreset AFT ages in the east are between 25 Myr and 30 Myr. Linear inversions of these synthetic data produce a spatial gradient in exhumation rates that is damped compared to the input gradient. The linear inversion of the data, using a 30-km spatial correlation length and an a priori erosion rate of 0.35 ± 0.1 mm yr^{-1}, predicts resolved moderate increases in exhumation rates throughout the model domain, with normalized differences varying from 0.08 in the east to 0.46 in the west (Extended Data Fig. 6e). As in the previous synthetic tests, reducing the correlation length to 10 km does not strongly change the pattern but reduces the resolution, in particular from 6–4 Myr ago (Extended Data Fig. 6f–h). When the a priori erosion rate is reduced to 0.1 ± 0.1 mm yr^{-1}, the model predicts strong decreases (normalized difference up to −0.42) in the east and strong increases (up to 0.61) in the west (Extended Data Fig. 6k). These synthetic tests illustrate again how samples with multiple thermochronometers do not counteract the spatial correlation bias. Moreover, this example shows the dependence of the linear inversion results on the assumed a priori erosion rate, as the model tends towards these rates where the resolution is relatively low.

The linear inversion model found spurious erosion-rate increases for all of the synthetic tests, which were assigned spatially variable but temporally constant exhumation rates. Overall, shorter spatial correlation lengths restrict the area for which spurious increases were interpreted, but they do not eliminate spurious increases, even in the synthetic test of the western Alps, where data are dense, multiple thermochronometer ages are available for several individual samples, and numerous steep elevation transects characterize the dataset.

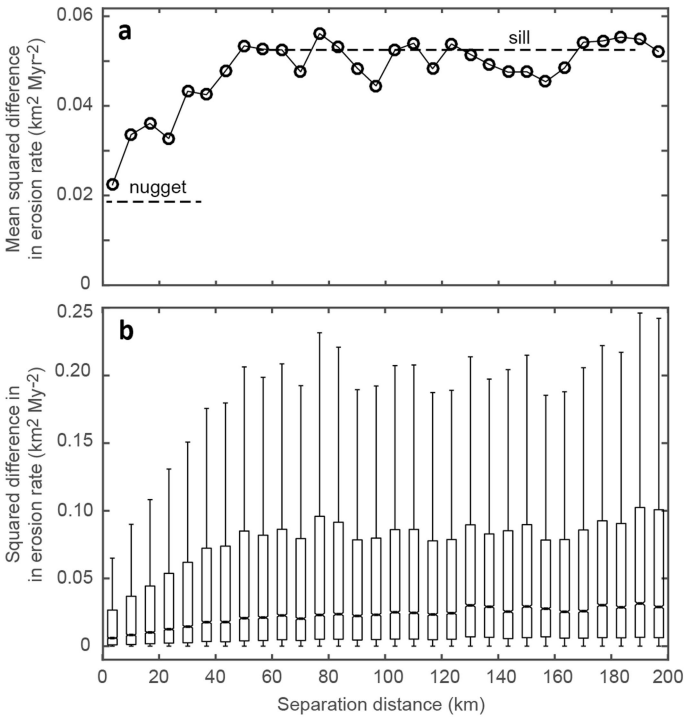
Our synthetic tests furthermore illustrate that choosing a higher resolution cut-off (higher than 0.25) reduces the regions for which spurious increases are resolved; however, even at resolutions of 0.5–0.6, spurious increases occur. On a global scale, increasing the resolution cut-off to 0.5 would result in only seven locations in the world with resolved late-Cenozoic increases in exhumation rates (the western European Alps, the Apennines, Taiwan, Fiordland, a single point in the Wasatch range and a single point in the central Himalaya), whereas a cut-off of 0.6 reduces the number of locations to three (the western Alps, the Apennines and Taiwan). Nevertheless, as we argued earlier and in Supplementary Information, most of those increases are spurious.

We also find that the magnitude of the acceleration is dependent on the chosen a priori erosion rate. In their discussion of the sensitivity of the inversion results to the a priori erosion rate, the authors of ref. [6] note that "choosing a prior erosion rate that is substantially different from the actual solution will lead to a wrong solution". However, they used a constant and relatively low (0.35 ± 0.1 mm yr^{-1}) a priori erosion rate for all 'tectonically active' regions, potentially further exacerbating the inferred recent increase in erosion rates.

Data availability. Data for the European Alps and New Zealand are based on the global thermochronology data compilation published in ref. [6]. All other thermochronology data shown were compiled by the authors, but can also be obtained from the supplementary information of ref. [6]. All shaded relief maps (for example, in Figs. 3 and 4 and in Extended Data Figs. 3, 5 and 6) were created in ArcGIS 10.3.1 from 3-arcsecond Shuttle Radar Topography Mission (SRTM) digital elevation data available from the US Geological Survey. Input files for Pecube that were used to predict data for our synthetic tests and input files for the linear inversion model are available at https://github.com/TaylorSchildgen/SpatialCorrelationBias. The linear inversion model is available at https://zenodo.org/record/1215685.

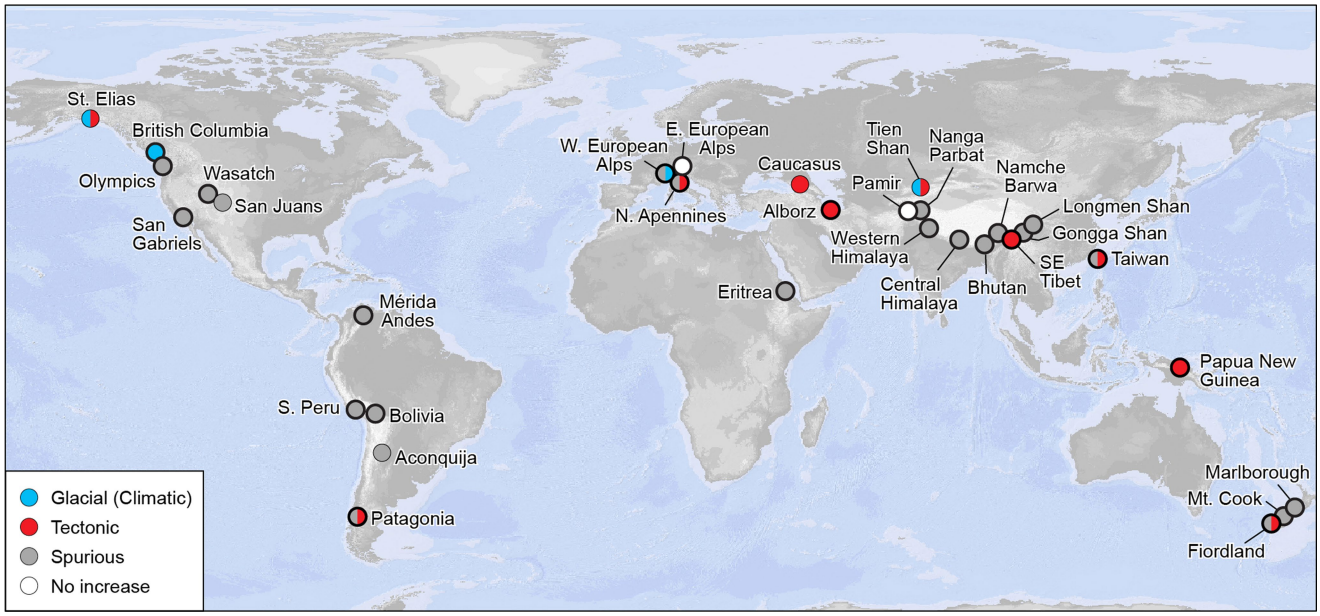
34. Brandon, M. T., Roden-Tice, M. K. & Garver, J. I. Late Cenozoic exhumation of the Cascadia accretionary wedge in the Olympic Mountains, northwest Washington State. *Geol. Soc. Am. Bull.* **110**, 985–1009 (1998).
35. Braun, J., van der Beek, P. & Batt, G. *Quantitative Thermochronology: Numerical Methods for the Interpretation of Thermochronological Data* (Cambridge Univ. Press, Cambridge, 2006).
36. Dodson, M. H. Closure temperature in cooling geochronological and petrological systems. *Contrib. Mineral. Petrol.* **40**, 259–274 (1973).
37. Laslett, G. M., Green, P. F. & Duddy, I. R. Thermal annealing of fission tracks in apatite 2. A quantitative analysis. *Chem. Geol. Isot. Geosci. Sect.* **65**, 1–13 (1987).
38. Mancktelow, N. S. & Grasemann, B. Time-dependent effects of heat advection and topography on cooling histories during erosion. *Tectonophysics* **270**, 167–195 (1997).
39. Braun, J. et al. Quantifying rates of landscape evolution and tectonic processes by thermochronology and numerical modeling of crustal heat transport using PECUBE. *Tectonophysics* **524–525**, 1–28 (2012).
40. Cressie, N. *Statistics for Spatial Data* (Wiley Interscience, Hoboken, 1993).
41. Farley, K. A. Helium diffusion from apatite: general behavior as illustrated by Durango fluorapatite. *J. Geophys. Res.* **105**, 2903–2914 (2000).
42. Reiners, P. W., Spell, T. L., Nicolescu, S. & Zanetti, K. A. Zircon (U-Th)/He thermochronometry: He diffusion and comparisons with $^{40}Ar/^{39}Ar$ dating. *Geochim. Cosmochim. Acta* **68**, 1857–1887 (2004).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



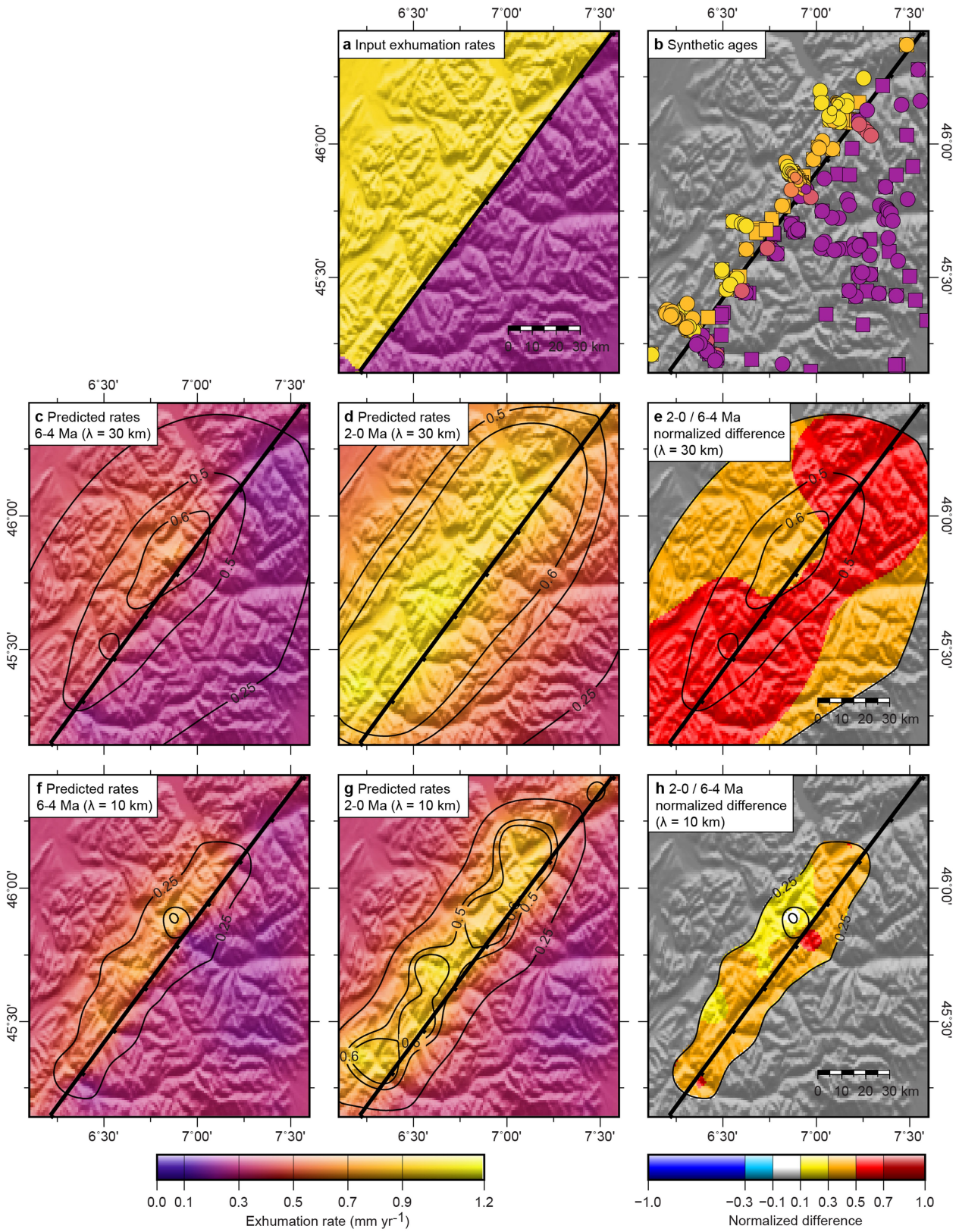
Extended Data Fig. 1 | Empirical semi-variogram of erosion rates derived from AFT data from the European Alps. a, Gamma plot, showing the mean of the squared differences (variance) in erosion rates for each distance bin. The plot shows a negative exponential pattern, as described in ref. [18]. The variance at a distance of zero, defined as the 'nugget', is non-negligible. The nugget is considered to represent the measurement error plus the variance at distances smaller than the smallest measurable distance[40]. The 'sill' is the relatively steady variance value reached at distances beyond the initial increase[40]. The nugget-to-sill ratio, which in this case is about 0.35, indicates that approximately 35% of the variance is unaccounted for with the spatial correlation function[40]. **b**, Box plot of the range of squared differences in erosion rates, with boxes outlining quartiles of data, and error bars corresponding to 1.5 times the inner quartile distance. Horizontal lines within boxes show medians of data within each distance bin.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



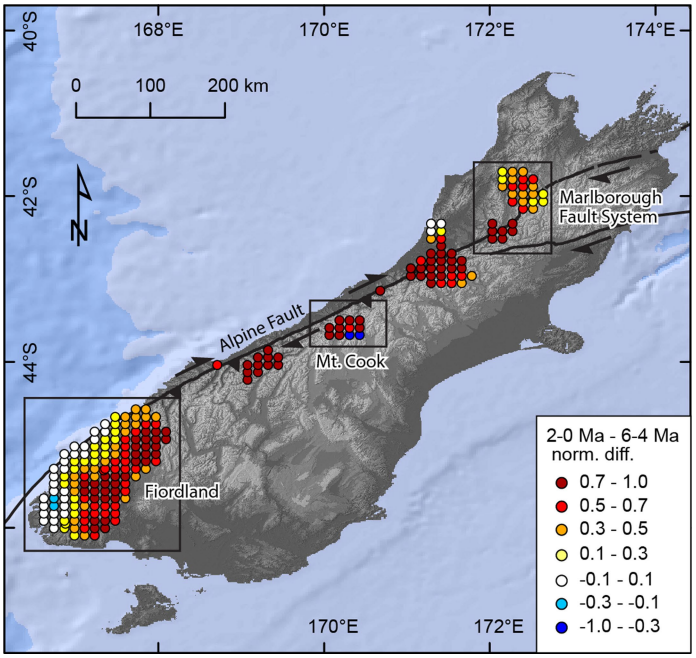
Extended Data Fig. 2 | Global coverage and inferred cause of resolved changes in late-Cenozoic erosion rates. The data are based on the inversion results of ref. [6]. Locations where erosion rates are resolved both from 2–0 Myr ago and from 6–4 Myr ago are shown with thin black circles, whereas locations where erosion rates are resolved in all four 2-Myr time bins between 8 Myr ago and the present are shown with thick black circles. Circles filled with multiple colours indicate contributions from multiple factors.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



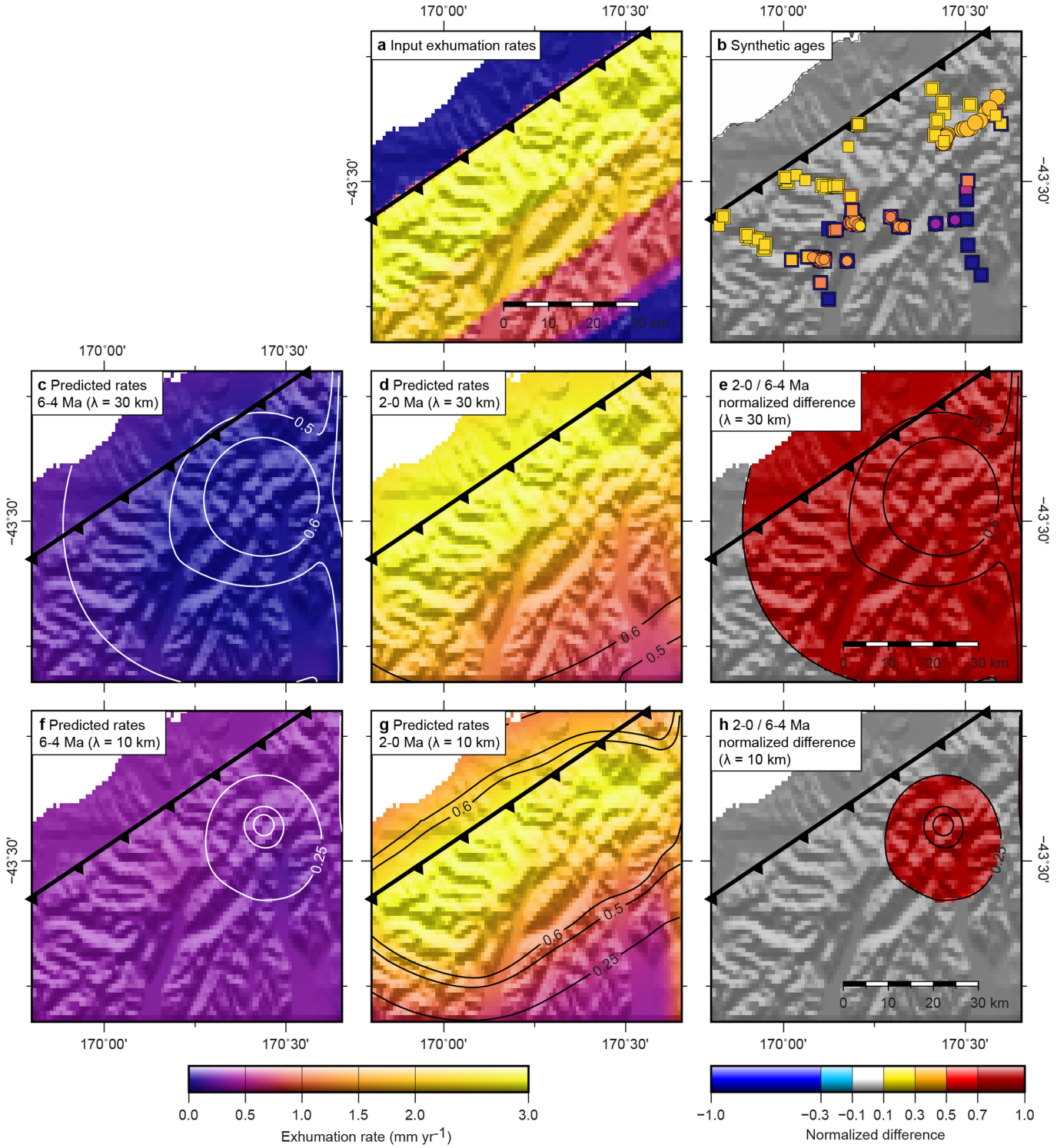
Extended Data Fig. 3 | Synthetic test for the western Alps. a, Input exhumation rates are 1 mm yr^{-1} in the external zone and 0.25 mm yr^{-1} in the internal zone; they are held constant for 20 Myr. **b**, Predicted thermochronological ages; different symbols for thermochronological systems and colour scale for ages are as in Fig. 3a. **c**–**e**, Predicted erosion rates for a correlation length (λ) of 30 km from 6–4 Myr ago (**c**), from 2–0 Myr ago (**d**), and normalized difference in erosion rates from the comparison of **c** to **d** (as defined in equation (1)), shown only where the resolution in each time bin is greater than 0.25 (**e**). Contours in **c**–**e** show the predicted resolution. **f**–**h**, Same as **c**–**e**, but for a correlation length of 10 km. Ma, million years ago.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



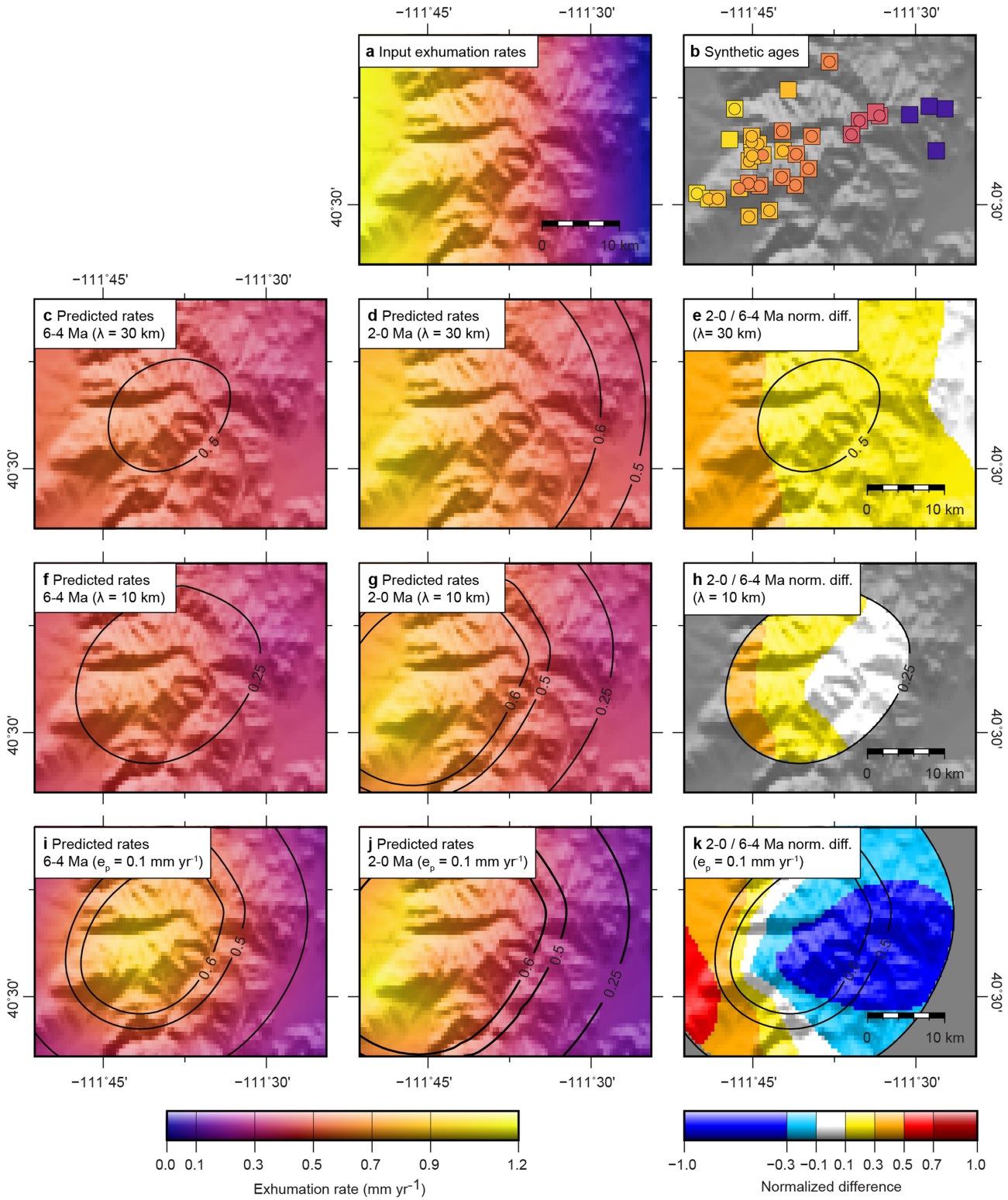
Extended Data Fig. 4 | Overview map of the South Island, New Zealand. Circles are locations where erosion rates are resolved (resolution >0.25) from both 2–0 Myr ago and from 6–4 Myr ago from the extended data of ref. [6]. Black boxes outline the regions described in the case studies of Mount Cook (Fig. 4), Fiordland (Supplementary Fig. 8) and Marlborough fault system (Supplementary Fig. 16).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 5 | Synthetic test for the Mount Cook region in New Zealand. a, Input exhumation rates are predicted from a kinematic model of overthrusting along a fault with a ramp–flat geometry[30]; rates vary from 2.2 mm yr^{-1} adjacent to the fault to zero far from it, and they are held constant for 10 Myr. **b**, Predicted thermochronological ages; different symbols for thermochronological systems and the colour scale for ages are as in Fig. 4a. **c**–**e**, Predicted erosion rates for a correlation length of 30 km from 6–4 Myr ago (**c**), from 2–0 Myr ago (**d**), and normalized difference in erosion rates from the comparison of **c** to **d** (as defined in equation (1)), shown only where the resolution in each time bin is greater than 0.25 (**e**). Contours in **c**–**e** show the predicted resolution. **f**–**h**, Same as plots **c**–**e**, but for a correlation length of 10 km.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 6 | Synthetic test for the Wasatch Mountains. **a**, Input exhumation rates increase smoothly from zero at the eastern boundary to 1.2 mm yr^{-1} at the western boundary and are held constant for 12 Myr. **b**, Predicted thermochronological ages. Symbols for thermochronological systems and colour scale for ages are as in Supplementary Fig. 1a. **c**–**e**, Predicted erosion rates for a correlation length of 30 km from 6–4 Myr ago (**c**), from 2–0 Myr ago (**d**), and normalized difference in erosion rates from the comparison of **c** to **d** (as defined in equation (1)), shown only where the resolution in each time bin is >0.25 (**e**). Contours in **c**–**e** show the predicted resolution. **f**–**h**, Same as **c**–**e**, but for a correlation length of 10 km. **i**–**k**, Same as **f**–**h**, but for a prior erosion rate (e_p) of 0.1 mm yr^{-1} instead of 0.35 mm yr^{-1}.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Global inversion results from ref. [6] and likely causes for erosion-rate increases

	Bins 6-4 Ma and 2-0 Ma		All four bins		Normalized difference 2-0 Ma minus 6-4 Ma					
Location	Points	% total	Points	% total	min	max	average	median	Tectonic setting	Cause of increase*
Alps	360	21.8	328	30.0	-0.44	0.93	0.31	0.30		
Eastern Alps	*70*	*4.2*	*69*	*6.3*	*0.00*	*0.13*	*0.07*	*0.08*	Post-orogenic	No significant increase
Western Alps	*290*	*17.6*	*259*	*23.7*	*-0.44*	*0.93*	*0.37*	*0.38*	Normal/thrust faulting	Spurious/Glacial
New Zealand	221	13.4	130	11.9	-0.67	0.99	0.51	0.57		
Marlborough Fault System	*28*	*1.7*	*14*	*1.3*	*0.15*	*0.99*	*0.53*	*0.49*	Transpression	Spurious
Mt. Cook region	*56*	*3.4*	*11*	*1.0*	*-0.67*	*0.98*	*0.70*	*0.83*	Accretionary wedge	Spurious
Fiordland	*137*	*8.3*	*105*	*9.6*	*-0.14*	*0.90*	*0.43*	*0.47*	Accretionary wedge	Spurious/Tectonic
W. Himalaya	200	12.1	122	11.2	-0.43	0.81	0.21	0.20	Thrust ramp	Spurious
Patagonia	146	8.8	60	5.5	-0.06	0.85	0.27	0.22	Transpression	Spurious/Tectonic
Northern Apennines	117	7.1	91	8.3	-0.14	0.99	0.43	0.46	Accretionary wedge	Spurious/Tectonic
Taiwan	87	5.3	27	2.5	-0.94	0.99	0.61	0.74	Accretionary wedge	Spurious/Tectonic
Central Himalaya	75	4.5	53	4.8	-0.83	0.91	0.20	0.10	Thrust ramp	Spurious
Namche Barwa	66	4.0	49	4.5	-0.95	0.89	0.29	0.35	Crustal-scale pop-up	Spurious
Bhutan	53	3.2	38	3.5	-0.29	0.48	0.08	0.09	Thrust ramp	Spurious
Gongga Shan	49	3.0	42	3.8	-0.09	0.98	0.41	0.28	Transpression	Spurious
Olympic Mts.	39	2.4	37	3.4	0.13	0.56	0.31	0.28	Accretionary wedge	Spurious
Merida Andes	36	2.2	6	0.5	0.32	0.79	0.61	0.63	Transpression	Spurious
British Columbia	33	2.0	21	1.9	0.09	0.61	0.34	0.29	Valley incision	Glacial
Nanga Parbat	24	1.5	13	1.2	0.52	0.91	0.74	0.75	Crustal-scale pop-up	Spurious
Longmen Shan	21	1.3	16	1.5	0.20	0.65	0.43	0.43	Thrust ramp	Spurious
San Gabriel Mts.	18	1.1	17	1.6	0.36	0.78	0.56	0.55	Transpression	Spurious
Eritrea	18	1.1	6	0.5	0.00	0.98	0.47	0.38	Normal-fault block	Spurious
Alborz	16	1.0	9	0.8	0.23	0.81	0.46	0.42	Accretionary wedge	Tectonic
Wasatch	14	0.8	9	0.8	0.05	0.29	0.20	0.21	Normal-fault block	Spurious
St. Elias	11	0.7	0	0.0	0.51	0.62	0.59	0.60	Accretionary wedge	Tectonic/Glacial
Papua New Guinea	10	0.6	3	0.3	0.00	0.80	0.15	0.03	Accretionary wedge	Tectonic
San Juan Mts.	7	0.4	0	0.0	0.59	0.99	0.83	0.85	Post-orogenic	Spurious
Pamir	7	0.4	6	0.5	0.02	0.08	0.05	0.05	Normal-fault block	No significant increase
Southern Peru	6	0.4	4	0.4	0.27	0.50	0.43	0.47	Valley incision	Spurious
Bolivia	5	0.3	5	0.5	0.12	0.28	0.18	0.15	Valley incision	Spurious
Southeast Tibet	3	0.2	2	0.2	0.05	0.12	0.07	0.05	Thrust ramp	Tectonic
Aconquija	3	0.2	0	0.0	0.48	0.52	0.50	0.50	Thrust block	Spurious
Greater Caucasus	2	0.1	0	0.0	0.56	0.65	0.60	0.60	Thrust ramp	Tectonic
Tien Shan	1	0.1	0	0.0	0.35	0.35	0.35	0.35	Thrust block	Tectonic/Glacial

*See Supplementary Information for detailed descriptions. Ma, million years ago.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 2 | Thermochronological ages corresponding to steady erosion rates

Erosion rate (mm yr^{-1})	AHe age (Ma)	AFT age (Ma)	ZHe age (Ma)	ZFT age (Ma)
0.07	31.3	50.5	85.7	104.0
0.10	22.4	35.8	60.4	73.1
0.15	15.2	24.1	40.4	48.7
0.20	11.5	18.1	30.2	36.3
0.30	7.6	11.9	19.8	23.8
0.40	5.7	8.7	14.6	17.4
0.50	4.4	6.8	11.4	13.6
0.60	3.6	5.6	9.2	11.0
0.70	3.0	4.6	7.7	9.2
0.80	2.6	3.9	6.5	7.7
0.90	2.2	3.4	5.6	6.7
1.20	1.5	2.3	3.9	4.6
1.50	1.1	1.7	2.8	3.4
2.00	0.7	1.1	1.8	2.2
3.00	0.4	0.6	1.0	1.1

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 3 | Diffusion parameters used for closure-temperature calculations

Thermochronometer (reference)	E_a (kJ mol^{-1})	D_0 (cm^2 s^{-1})	a (μm)	T_c; $\dot{e}$=0.1 mm yr^{-1} (°C)†	T_c; $\dot{e}$=1.0 mm yr^{-1} (°C)
AHe [41]	138	50	100	69	88
ZHe [42]	169	0.46	100	189	219
AFT [14]‡	147	3.73×10^4		110	134
ZFT [14]§	208	1.82×10^6		230	258

In all calculations, $R = 8.3145$ J K^{-1} mol^{-1}, $A = 27$, $L = 30$ km, $\kappa = 30$ km^2 Myr^{-1}, and the initial geothermal gradient is 30 °C km^{-1} ($T_0 = 0$ °C; $T_L = 900$ °C).
†Predicted closure temperature (T_c) for slow (0.1 mm yr^{-1}) and rapid (1 mm yr^{-1}) exhumation rates ($\dot{e}$).
‡Values corresponding to the 'average apatite composition' of ref. [14], with a value of D_0/a^2 (s^{-1}).
§Values corresponding to the 'natural, radiation damaged' zircon of ref. [14], with a value of D_0/a^2 (s^{-1}).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

https://doi.org/10.1038/s41586-018-0264-2

Low-temperature crystallization of granites and the implications for crustal magmatism

Michael R. Ackerson[1,2]*, B. O. Mysen[1], N. D. Tailby[3] & E. B. Watson[2]

The structure and composition of granites provide clues to the nature of silicic volcanism, the formation of continents, and the rheological and thermal properties of the Earth's upper crust as far back as the Hadean eon during the nascent stages of the planet's formation[1–4]. The temperature of granite crystallization underpins our thinking about many of these phenomena, but evidence is emerging that this temperature may not be well constrained. The prevailing paradigm holds that granitic mineral assemblages crystallize entirely at or above about 650–700 degrees Celsius[5–7]. The granitoids of the Tuolumne Intrusive Suite in California tell a different story. Here we show that quartz crystals in Tuolumne samples record crystallization temperatures of 474–561 degrees Celsius. Titanium-in-quartz thermobarometry and diffusion modelling of titanium concentrations in quartz indicate that a sizeable proportion of the mineral assemblage of granitic rocks (for example, more than 80 per cent of the quartz) crystallizes about 100–200 degrees Celsius below the accepted solidus. This has widespread implications. Traditional models of magma formation require high-temperature magma bodies, but new data[8,9] suggest that volcanic rocks spend most of their existence at low temperatures; because granites are the intrusive complements of volcanic rocks, our downward revision of granite crystallization temperatures supports the observations of cold magma storage. It also affects the link between volcanoes, ore deposits and granites: ore bodies are fed by the release of fluids from granites below them in the crustal column; thus, if granitic fluids are hundreds of degrees cooler than previously thought, this has implications for research on porphyry ore deposits. Geophysical interpretations of the thermal structure of the crust and the temperature of active magmatic systems will also be affected.

Terrestrial granitoids (coarsely crystalline igneous rocks whose compositions range from 65 wt% to 77 wt% SiO_2) are the ultimate products of the differentiation of our planet, and the processes leading to these highly evolved rocks are integral to the growth and maturation of the continental crust. As far as we know, these processes are also unique to the Earth. Knowledge of the conditions under which granitic rocks crystallize is critical to our understanding of global-scale phenomena including the formation and stabilization of continents on the early Earth[4], the petrogenetic link between plutonic and volcanic rocks, the events leading to large-volume explosive volcanic eruptions[10] and interpretations of seismic data from active volcanic regions[11]. Most experimental studies to determine the solidification temperature (wet solidus) of granitic bulk compositions have led to the view that granitic (and subsequently all silicic magmatic rocks) will crystallize almost completely when cooled to 650–700 °C[5–7]. However, many of the minerals in granitic rocks record temperatures well below 650 °C[12–14]. Constrained by the assumption that minerals in granites cannot crystallize below 650 °C, and in lieu of alternative explanations, these low temperatures have commonly been interpreted as subsolidus re-equilibration during metamorphism or hydrothermal alteration.

In contrast to the accepted paradigm, several experimental studies and observations of natural samples indicate that crystallization of granitic mineral assemblages is possible below the generally accepted wet solidus. For example, quartz and feldspars can crystallize in equilibrium with low-volume (around 1%–5%) interstitial hydrous peralkaline [(Na + K) > Al] melts along a cotectic curve to temperatures below 400 °C in alkali-aluminosilicate systems[7,15,16], and andesitic bulk compositions placed in thermal gradients can crystallize granitic mineral assemblages at temperatures as low as 350 °C[17] from hydrous peralkaline melts (>40 wt% H_2O). It is unclear whether these processes can occur on the scale of batholiths, but such experiments highlight the potential for quartz and feldspars to crystallize at temperatures hundreds of degrees below those previously thought possible.

Here, we use Ti-in-quartz thermobarometry in conjunction with diffusion modelling and cathodoluminescence imaging to show that quartz in granitic rocks from the Tuolumne Intrusive Suite (TIS)—an archetype of modern subduction-related continental crust formation—crystallized at temperatures hundreds of degrees below the traditionally accepted granitic wet solidus.

Titanium concentrations of 20–40 p.p.m. on the rims of quartz crystals from the TIS indicate crystallization at temperatures of about 474–561 °C, 120–230 °C below the generally accepted granodiorite wet solidus[6] of about 690 °C at 2 kbar pressure. Assuming that the rims of quartz crystals in granitoids of the TIS represent late-stage near-solidus crystallization, then the Ti content near the rims of quartz crystals can be predicted by linking the granodiorite wet solidus (a curve in temperature–pressure (*T*–*P*) space)[6] with a model for Ti concentration in quartz that varies as a function of *T*, *P* and the Ti activity relative to rutile saturation in the melt, $a_{TiO_2}^{melt-rutile}$ (ref. [18]). For a pressure range of 1.6–2.4 kbar and $a_{TiO_2}^{melt-rutile} = 0.5–0.6$ (see Methods for discussion of *P* and $a_{TiO_2}^{melt-rutile}$), Ti contents from about 132 p.p.m. to 219 p.p.m. would be predicted if the rims of quartz crystallized at the granodiorite wet solidus (Fig. 1). This expected concentration is considerably higher than the average rim concentrations of Ti (20–40 p.p.m.) observed throughout the TIS (Fig. 2).

The low Ti contents of quartz rims from the TIS demonstrate that these granitoids experienced at least late-stage crystallization at temperatures below the traditional granodiorite wet solidus. However, if we assume a uniform $a_{TiO_2}^{melt-rutile}$ of 0.5 or greater throughout quartz crystallization—a reasonable estimate given the ubiquitous presence of titanite throughout the TIS—then the main volume of quartz within these granitic rocks crystallized below the expected wet solidus temperatures. Low-temperature quartz crystallization is consistent with other independent temperature estimates, including temperatures calculated based on the compositions of feldspar assemblages (482–519 °C for the Cathedral Peak granodiorite)[12], oxygen isotope fractionation between zircons and quartz (average *T* for the entire TIS of about 587 °C, with measurements as low as 521 °C)[19] and amphibole chemistry (approximately 300–450 °C)[13] from the TIS (Extended Data Fig. 8).

[1]Geophysical Laboratory, Carnegie Institution of Washington, Washington, DC, USA. [2]Department of Earth and Environmental Sciences, Rensselaer Polytechnic Institute, Troy, NY, USA. [3]Department of Earth and Planetary Sciences, American Museum of Natural History, New York, NY, USA. *e-mail: ackerm4@rpi.edu

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

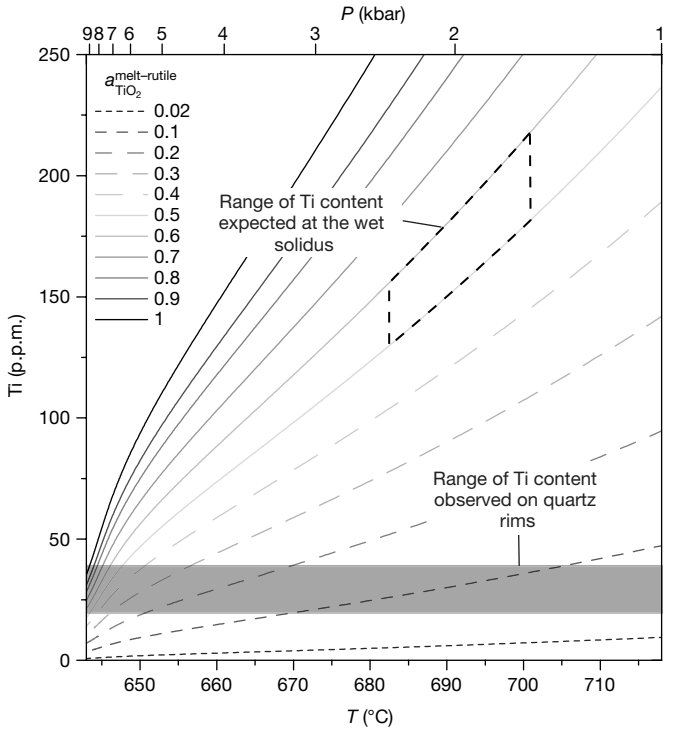


Fig. 1 | Expected and observed quartz crystallization temperatures at the granodiorite wet solidus. The granodiorite wet solidus (sample JSP 6-2)[6] is expressed as expected Ti content in quartz (Ti in p.p.m.) at a given solidus P and T. Curves represent the solidus at different $a_{TiO_2}^{melt-rutile}$. A granitic $a_{TiO_2}^{melt-rutile} = 0.5–0.6$ for rutile-absent rocks of the TIS at 1.6–2.4 kbar (0.16–0.24 GPa) pressure (see text for details) yields estimated solidus Ti-in-quartz concentrations from about 132 p.p.m. to 219 p.p.m. (dashed polygon). Measured Ti concentrations near the rims of TIS quartz average between 20 p.p.m. and 40 p.p.m. (grey-shaded rectangle), below the estimated Ti content, suggesting that the quartz crystals grew at temperatures up to about 227 °C below the solidus.

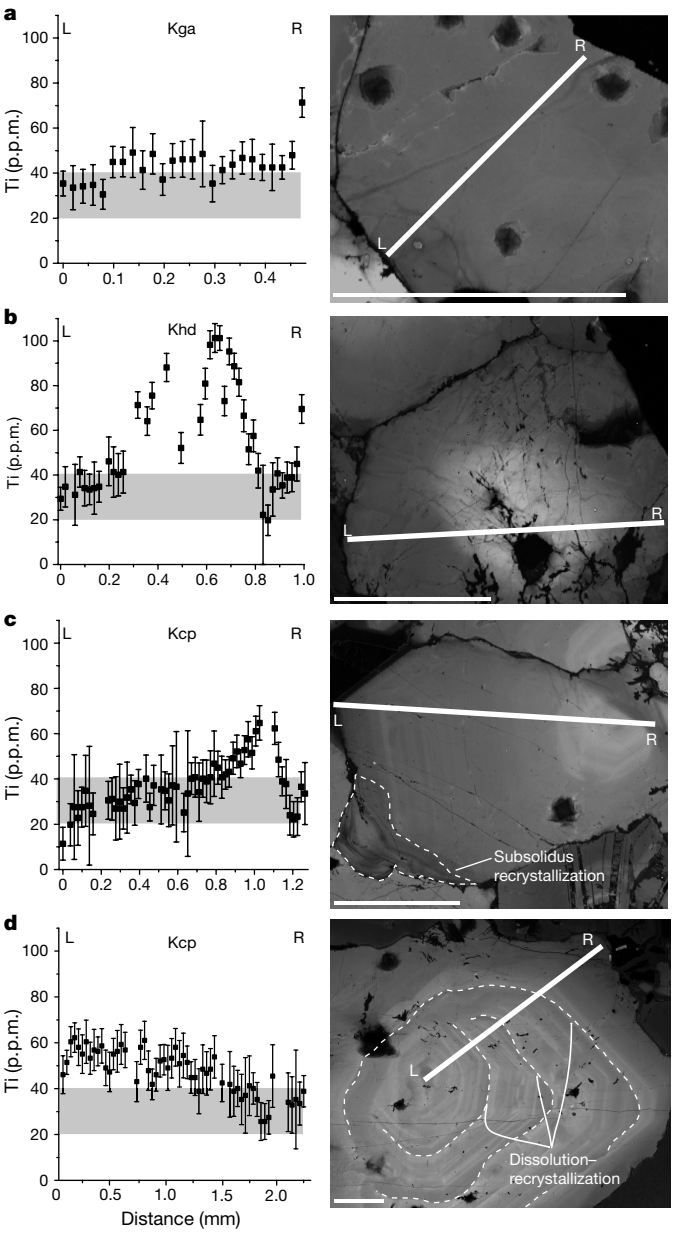


Fig. 2 | Titanium concentrations along transects through quartz crystals from granitoids from the Tuolumne Intrusive Suite (TIS), and corresponding cathodoluminescence maps. a, Glen Aulin tonalite (Kga); **b**, porphyritic Half Dome granodiorite (Khd); **c**, **d**, Cathedral Peak granodiorite (Kcp). Cathodoluminescence images are filtered for the blue wavelengths 445–455 nm. Blue cathodoluminescence intensity is directly proportional to the amount of Ti in the quartz lattice. Error bars are 1σ, and scale bars are 500 μm. The grey bar represents the average rim composition of TIS quartz (20–40 p.p.m. Ti). The black holes in several cathodoluminescence images are pits formed during laser-ablation inductively coupled plasma mass spectrometry.

Because $a_{TiO_2}^{melt-rutile}$ is not precisely constrained for these rocks, and also because an alternative calibration of the Ti-in-quartz thermobarometer does exist[18,20] (see Methods for further discussion), we sought additional constraints on the crystallization temperature of the TIS quartz rims. Given that the Arrhenius relation for Ti diffusion in quartz is well characterized[18], the extent of diffusive relaxation of Ti concentration profiles can be used to place constraints on time–temperature histories of host crystals. Accordingly, to further substantiate the occurrence of low-T crystallization recorded by Ti content, we ran forward numerical simulations of Ti diffusion with the goal of matching Ti concentration profiles obtained from cathodoluminescence imaging of quartz crystals from several units of the TIS. This technique provides an independent assessment of quartz crystallization temperature because diffusion is insensitive to variables that are important in Ti-in-quartz thermometry (pressure and $a_{TiO_2}^{melt-rutile}$).

The intensity of light at a wavelength of approximately 452 nm in the cathodoluminescence spectrum for quartz is directly proportional to the Ti content of quartz. Variations in intensity of the 452-nm peak can be used as a proxy for Ti concentration gradients in quartz[21] and can subsequently be used as observable constraints to model diffusive relaxation of Ti concentration profiles. The high spatial resolution of cathodoluminescence chemical profiles is imperative for matching small-scale Ti concentration gradients in quartz with diffusion model results. For a cooling rate of 36.36 °C per million years (Myr)[22], observed concentration gradients match calculated diffusion profiles only when the assumed initial temperature is at or below 550 °C (Fig. 3b). Even at unrealistically high cooling rates (for example, 300 °C Myr^{-1}), diffusion models do not match the observed (very steep) concentration profiles if the initial temperature is set at the canonical solidus temperature of 700 °C (Fig. 3c). Simply stated, the observed Ti concentration profiles in TIS quartz grains are far too steep to have originated at temperatures as high as 650–700 °C, even if cooling of the pluton is assumed to have been implausibly fast. Diffusion modelling thus confirms the low temperatures calculated using Ti-in-quartz thermobarometry. Furthermore, diffusion modelling of Ti concentration gradients in quartz crystals from the Cathedral Peak granodiorite (Fig. 3; Extended Data Fig. 5) indicates that about 90% of the quartz (not just the rims) crystallized at low temperatures. Combined with the compositions of feldspars from this unit, this indicates that quartz and the feldspars[12] (which account for about 95% of the Cathedral Peak granodiorite[23])

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

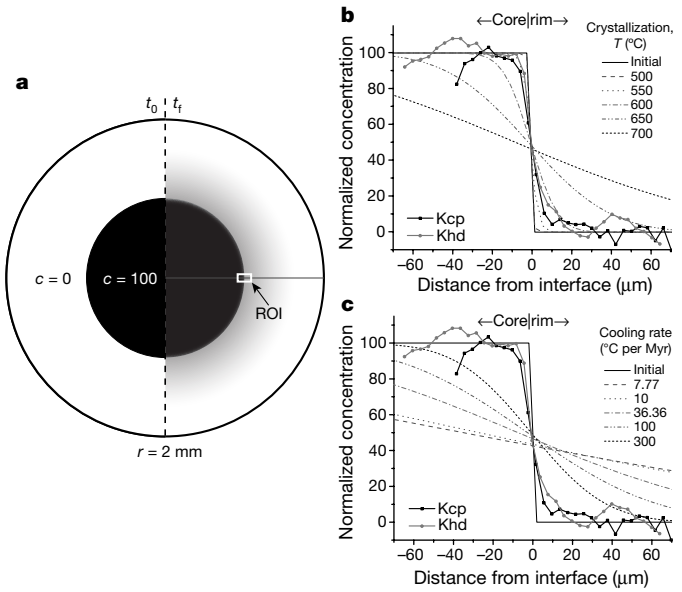


Fig. 3 | Schematic model and calculations of diffusive relaxation of Ti step profiles in quartz compared with measured profiles in natural crystals. a, In the model, a spherical quartz crystal with an initial central concentration *c* of 100 p.p.m. Ti and outer concentration of 0 p.p.m. Ti at time t_0 is allowed to diffusively relax at a range of initial temperatures and cooling rates, using the experimentally determined diffusion data for Ti in quartz[29]. The model profiles were obtained by numerical integration of the non-steady-state diffusion equation for a spherical geometry[30]. The outer edge of the crystal was assumed to be a zero-flux boundary, but qualitative outcomes are not sensitive to this assumption. ROI, region of interest. **b**, Results of diffusion modelling at a range of initial temperatures cooled at a constant 36.36 °C Myr^{-1} compared with normalized blue-cathodoluminescence intensity profiles from the Kcp and Khd demonstrate that low crystallization temperatures are required to generate diffusion profiles like those observed in the TIS. **c**, Diffusion profiles calculated by cooling quartz from the wet solidus (about 690–700 °C at 2 kbar) at a range of reasonable cooling rates demonstrate that even for very fast cooling rates, near-solidus crystallization temperatures are too hot to preserve the zoning observed in quartz from the TIS. The cooling rate of 7.77 °C Myr^{-1} was calculated by assuming linear cooling from the solidus at about 700 °C to 0 °C, over a time of 90 Myr. The cooling rate of 36.36 °C Myr^{-1} was estimated from the closure temperatures and age estimates for minerals (from zircon age and closure temperature to biotite age and closure temperature) from the McDoogle Quartz Monzodiorite[22].

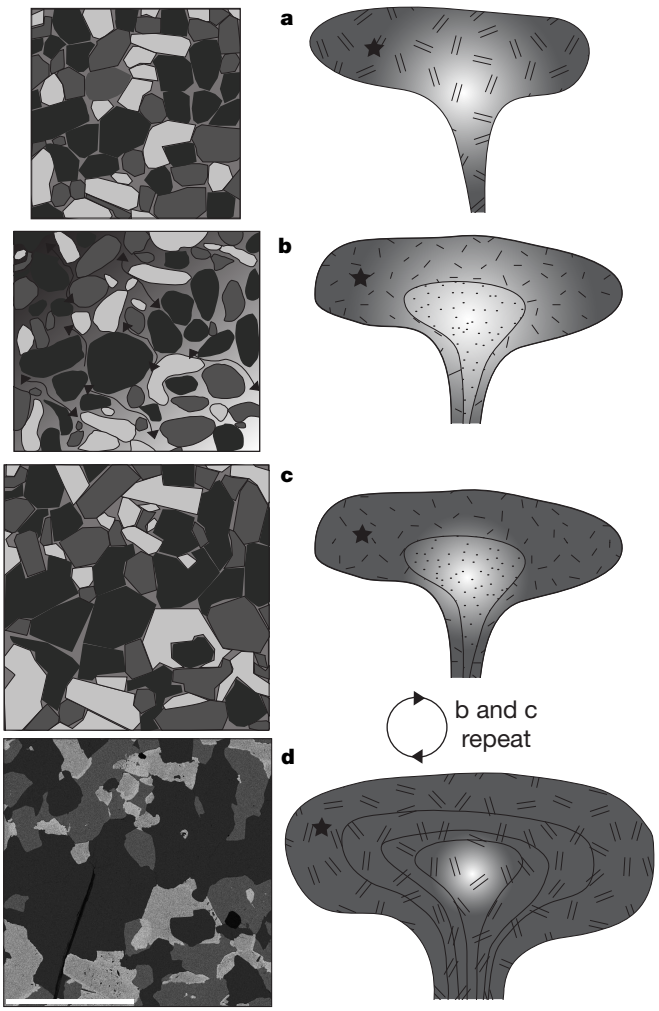


Fig. 4 | Schematic model of low-*T* crystallization in the TIS at hand-sample and pluton scales. Left panel is hand-sample view of a region of interest (star on right panel) in the TIS with quartz (black), K-spar (grey) and plagioclase (light grey). Right panel is the pluton-scale perspective (brighter shades are higher temperatures). **a**, A crystal-rich magma with interstitial melt cooled to near 500 °C. **b**, A new pulse of magma (temperatures of the new magma are likely to be above 700 °C) brings heat and imposes a new thermal gradient on the system. The thermal gradient promotes generation and movement of hydrous melts and dissolution of minerals in the pre-existing rocks at temperatures below 700 °C. The amount of melt present at low temperature will be about 5%. **c**, Cooling and crystallization of minerals from the melt. As magmatism continues, **b** and **c** repeat. **d**, Region of interest is no longer in thermal or chemical contact with new magmas, or magmatism ceases, and the system cools and crystallizes completely. The volume of interest is an open system with thermal pulses both dissolving the minerals within it and bringing new material to the region across thermal gradients, effectively increasing the amount of material within. This mass addition is signified by the increase in area of the left-hand side of the figure. The amount of low-*T* quartz and feldspar is in excess of 90%. Combined with the low volume of interstitial hydrous peralkaline melt (5%), this indicates that considerable recrystallization of the pre-existing quartz and feldspar in combination with mass added during thermal pulsing must occur. The left-hand side of **d** is a false-colour energy-dispersive spectroscopy (EDS) map of the Johnson Granite Porphyry (Si is black, Na is grey and K is light grey). Scale bar, 1 mm; all left-hand panels to same scale.

crystallized at temperatures as low as 500 °C. Despite limitations in the application of diffusion modelling to all quartz samples within the TIS (see Methods for discussion), the observed Ti-in-quartz temperatures, diffusion modelling and two-feldspar temperatures across the TIS (Extended Data Figs. 5 and 7) indicate that around 90% of the quartz and feldspars crystallized below the traditional wet solidus.

Development of a model for low-temperature crystallization in granites requires understanding the composition of the melts from which the minerals are crystallizing. A limited number of experimental studies on low-*T* mineral crystallization in granitic systems exist. The currently available data sets indicate that granitic crystallization near 500 °C requires moderately peralkaline [(Na + K) > Al] melts, which could be either silicate-laden aqueous fluid or water-rich silicate melt[7,15]. An additional requirement is a thermal gradient over which the melt is created and sustained[17], aided by long-lived, thermally fluctuating magmatism[24]. All of these conditions can be met within the parameters of the incremental emplacement model for the TIS, which states that large-volume, compositionally zoned plutons like the TIS form through protracted emplacement of small volumes of magma into the crust[10]. Small proportions of interstitial hydrous peralkaline melts that crystallize granitic mineral assemblages have been shown to produce metaluminous bulk compositions similar to the TIS[17], and evidence for thermal cycling is recorded in the dissolution–recrystallization features seen in K-feldspar megacrysts[12,25] and quartz (Fig. 2d) of the TIS.

If peralkaline melts are necessary for low-*T* crystallization, then the amount of low-temperature melt required at any given time will be

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

about 5%, such that the bulk of the system will remain metaluminous in composition despite the peralkaline nature of the melt. The solubility of peralkaline silicates in low-T hydrothermal fluids means that residual traces of these melts would probably be lost to low-grade alteration by metamorphic fluids. If future research reveals alternative melt compositions stable at low temperatures, the amount of melt present at 500 °C could reach as high as the rigid percolation threshold at about 50% crystallinity[26].

In conjunction with the low temperatures recorded in the quartz, the presence of dissolution–recrystallization features in quartz (Fig. 2) and feldspars[12] suggests that thermal fluctuations from incremental emplacement play an important role in crystallizing the TIS. Repeated injections of magma into the centre of the growing TIS will result in warming–cooling cycles in older units which can drive their dissolution and recrystallization (Fig. 4). Dissolution in the presence of thermal gradients could generate and sustain small-volume interstitial peralkaline water-rich melts that are stable at 500 °C.

The stability of granite-forming melts at low T can help to explain recent observations that eruptible magmas spend most of their lives at temperatures below the generally accepted wet solidus[8] and can help to bridge the gap between high-T volcanic systems and low-T hydrothermal porphyry precious-metal deposits[1]. In conjunction with feldspar temperature estimates from other studies, this also calls into question the model of subsolidus re-equilibration of granitic feldspars[27]. Lower crystallization temperatures will cause orders-of-magnitude increases in timescale estimates from quartz-based diffusion model estimates of magma residence times[28]. They will also influence interpretations of seismic tomography and magnetotelluric observations under active magmatic systems, which in turn can change geochemical models of magma accumulation rates and interpretations of the activity/quiescence of modern magmatic systems[2]. Low-temperature crystallization of granites will also lower estimates of crustal heat budgets, which directly influence models of crustal rheology[3], and could provide new geological models for the development of Hadean- and Archaean-aged continental crust[4].

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0264-2.

Received: 20 July 2017; Accepted: 23 April 2018;
Published online 27 June 2018.

1. Buret, Y. et al. From a long-lived upper-crustal magma chamber to rapid porphyry copper emplacement: reading the geochemistry of zircon crystals at Bajo de la Alumbrera (NW Argentina). *Earth Planet. Sci. Lett.* **450**, 120–131 (2016).
2. Comeau, M. J., Unsworth, M. J. & Cordell, D. New constraints on the magma distribution and composition beneath Volcán Uturuncu and the southern Bolivian Altiplano from magnetotelluric data. *Geosphere* **12**, 1391–1421 (2016).
3. Hacker, B. R., Kelemen, P. B. & Behn, M. D. Continental lower crust. *Annu. Rev. Earth Planet. Sci.* **43**, 167–205 (2015).
4. Reimink, J. R., Chacko, T., Stern, R. A. & Heaman, L. M. Earth's earliest evolved crust generated in an Iceland-like setting. *Nat. Geosci.* **7**, 529–533 (2014).
5. Luth, W. C., Jahns, R. H. & Tuttle, O. F. The granite system at pressures of 4 to 10 kilobars. *J. Geophys. Res.* **69**, 759–773 (1964).
6. Piwinskii, A. J. Experimental studies of granitoids from the Central and Southern Coast Ranges, California. *Tschermaks Mineral. Petrogr. Mitt.* **20**, 107–130 (1973).
7. Tuttle, O. F. & Bowen, N. L. Origin of granite in the light of experimental studies in the system $NaAlSi_3O_8$–$KAlSi_3O_8$–SiO_2–H_2O. *Geol. Soc. Am.* **74**, 1–146 (1958).
8. Rubin, A. E. et al. Rapid cooling and cold storage in a silicic magma reservoir recorded in individual crystals. *Science* **356**, 1154–1156 (2017).
9. Andersen, N. L., Jicha, B. R., Singer, B. S. & Hildreth, W. Incremental heating of Bishop Tuff sanidine reveals preeruptive radiogenic Ar and rapid remobilization from cold storage. *Proc. Natl Acad. Sci. USA* **114**, 12407–12412 (2017).
10. Coleman, D. S., Gray, W. & Glazner, A. F. Rethinking the emplacement and evolution of zoned plutons: geochronologic evidence for incremental assembly of the Tuolumne Intrusive Suite. *Calif. Geol.* **32**, 433–436 (2004).
11. Huang, H.-H. et al. The Yellowstone magmatic system from the mantle plume to the upper crust. *Science* **348**, 773–776 (2015).
12. Glazner, A. F. & Johnson, B. R. Late crystallization of K-feldspar and the paradox of megacrystic granites. *Contrib. Mineral. Petrol.* **166**, 777–799 (2013).
13. Challener, S. C. & Glazner, A. F. Igneous or metamorphic? Hornblende phenocrysts as greenschist facies reaction cells in the Half Dome Granodiorite, California. *Am. Mineral.* **102**, 436–444 (2017).
14. Gray, W., Glazner, A. F., Coleman, D. S. & Bartley, J. M. Long-term geochemical variability of the Late Cretaceous Tuolumne Intrusive Suite, central Sierra Nevada, California. *Geol. Soc. Lond. Spec. Publ.* **304**, 183–201 (2008).
15. Schairer, J. F. & Bowen, N. L. Melting relations in the systems Na_2O–Al_2O_3–SiO_2 and K_2O–Al_2O_3–SiO_2. *Am. J. Sci.* **245**, 193–204 (1947).
16. Tuttle, O. F. & Friedman, I. I. Liquid immiscibility in the system H_2O–Na_2O–SiO_2. *J. Am. Chem. Soc.* **70**, 919–926 (1948).
17. Huang, F. et al. Chemical and isotopic fractionation of wet andesite in a temperature gradient: experiments and models suggesting a new mechanism of magma differentiation. *Geochim. Cosmochim. Acta* **73**, 729–749 (2009).
18. Thomas, J. B. et al. TitaniQ under pressure: the effect of pressure and temperature on the solubility of Ti in quartz. *Contrib. Mineral. Petrol.* **160**, 743–759 (2010).
19. Lackey, J. S., Valley, J. W., Chen, J. H. & Stockli, D. F. dynamic magma systems, crustal recycling, and alteration in the Central Sierra Nevada Batholith: the oxygen isotope record. *J. Petrol.* **49**, 1397–1426 (2008).
20. Huang, R. & Audétat, A. The titanium-in-quartz (TitaniQ) thermobarometer: a critical examination and re-calibration. *Geochim. Cosmochim. Acta* **84**, 75–89 (2012).
21. Spear, F. S. & Wark, D. A. Cathodoluminescence imaging and titanium thermometry in metamorphic quartz. *J. Metamorph. Geol.* **27**, 187–205 (2009).
22. Davis, J. W., Coleman, D. S., Gracely, J. T., Gaschnig, R. & Stearns, M. Magma accumulation rates and thermal histories of plutons of the Sierra Nevada batholith, CA. *Contrib. Mineral. Petrol.* **163**, 449–465 (2012).
23. Bateman, P. C. & Chappell, B. W. Crystallization, fractionation, and solidification of the Tuolumne Intrusive Series, Yosemite National Park, California. *Geol. Soc. Am. Bull.* **90**, 465–482 (1979).
24. Lundstrom, C. C. The role of thermal migration and low-temperature melt in granitoid formation: can granite form without rhyolitic melt? *Int. Geol. Rev.* **58**, 371–388 (2016).
25. Moore, J. G. & Sisson, T. W. Igneous phenocrystic origin of K-feldspar megacrysts in granitic rocks from the Sierra Nevada batholith. *Geosphere* **4**, 387–400 (2008).
26. Vigneresse, J. L., Barbey, P. & Cuney, M. Rheological transitions during partial melting and crystallization with application to felsic magma segregation and transfer. *J. Petrol.* **37**, 1579–1600 (1996).
27. Plümper, O. & Putnis, A. The complex hydrothermal history of granitic rocks: multiple feldspar replacement reactions under subsolidus conditions. *J. Petrol.* **50**, 967–987 (2009).
28. Seitz, S. et al. Short magmatic residence times of quartz phenocrysts in Patagonian rhyolites associated with Gondwana breakup. *Geology* **44**, 67–70 (2016).
29. Cherniak, D. J., Watson, E. B. & Wark, D. A. Ti diffusion in quartz. *Chem. Geol.* **236**, 65–74 (2007).
30. Crank, J. *The Mathematics of Diffusion* (Clarendon, New York, 1979).

Acknowledgements All data for this manuscript are reported in the Extended Data. This research was supported by the Carnegie Institution of Washington's Postdoctoral Fellowship Program and the NASA Astrobiology Institute (grant NNA09DA80A) to the New York Center for Astrobiology. We thank N. Ackerson, M. J. Ackerson and V. Ackerson for assistance in the field, E. Bullock and J. Armstrong for their technical assistance in analysing quartz, and D. Coleman, J. Reimink, L. Wagner, H. Le Mével, S. Shirey and P. Ulmer for conversations during manuscript preparation.

Author contributions M.R.A. contributed the main ideas behind this study, analyses and writing of the manuscript. B.O.M., N.D.T. and E.B.W. provided critical expertise in the development of the proposed models. E.B.W. also contributed the diffusion model code.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0264-2.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to M.R.A.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Geological setting. The TIS is a group of relatively undeformed metaluminous granitoids within the Sierra Nevada Batholith (SNB). It is compositionally zoned from an older tonalite (Glen Aulin tonalite, Kga) near the exterior to the equigranular and porphyritic Half Dome granodiorites (Khd) and K-feldspar megacryst-bearing Cathedral Peak granodiorite (Kcp), to a younger granitic porphyry in the centre (Johnson granite porphyry, Kjp). The TIS crystallized over a period of about 10 Myr from 95–85 Myr ago[10]. This protracted crystallization window has led to the idea that large-volume, compositionally zoned plutons form through protracted incremental emplacement of small volumes of magma into the crust as opposed to rapid cooling of a single magma chamber[10]. Thermal cycling during repeated magma injections is one consequence of incremental emplacement[12]. Evidence for thermal cycling can be found in the dissolution–recrystallization features seen minerals from the TIS, most notably in K-feldspar megacrysts of the Kcp[12,25].

Samples for this study were collected from every major unit of the TIS (Extended Data Table 1, Extended Data Fig. 1), including aplites hosted in the Khd. The TIS also contains abundant aplite dykes that are interpreted as late-stage magmas from the host plutons[14]. Al-in-hornblende pressure estimates[16] suggest that the TIS crystallized at pressures of 1–3 kbar with most estimates[14,31,32] averaging around 2 kbar.

Analytical methods and numerical modelling. Samples of the TIS prepared for cathodoluminescence and electron microprobe analyses were mounted in 1″ epoxy rounds and sectioned with a diamond saw, polished to 1 μm with alumina polishing paste and placed in an ultrasonic distilled water bath to remove any polishing paste contamination. Samples were finally polished using colloidal silica.

Cathodoluminescence images of quartz crystals were collected using the xCLent IV cathodoluminescence detector on the JEOL 8530F at the Geophysical Laboratory of the Carnegie Institution. The intensity of the blue cathodoluminescence region of the quartz cathodoluminescence spectrum is directly proportional to the Ti content of the quartz[33,34]. The cathodoluminescence images of quartz crystals filtered for the blue region can therefore be used to qualitatively observe relative changes in the Ti content of individual quartz crystals.

Quantitative analyses were performed on the JEOL 8530F hyperprobe at the Geophysical Laboratory at 15 keV and 200 nA with a 10-μm spot. Titanium was measured on three pentaerythritol (PET) crystals with peak counting times of 320 s, yielding a detection limit for Ti of 6–8 p.p.m. Aluminium and silicon were measured on thallium acid phthalate (TAP) crystals for 320 s and 30 s, respectively. The accuracy of the Ti measurements was evaluated by measuring an internal standard Ti-bearing synthetic quartz crystal (sample AlQ10) whose Ti content has been independently evaluated with the Cameca SX-100 electron microprobe at Rensselaer Polytechnic Institute, yielding averages of (92 ± 10) p.p.m. Ti from the Geophysical Laboratory and (91 ± 14) p.p.m. Ti from Rensselaer Polytechnic Institute. Rim-to-rim and/or centre-to-rim transects of quartz crystals were performed on quartz from all units of the TIS. Rim analyses were primarily undertaken adjacent to feldspars to avoid secondary fluorescence effects from Ti-rich minerals. Predicted Ti content of quartz at the granodiorite wet solidus was calculated using the ref. [18] calibration for Ti-in-quartz and the experimentally derived granodiorite wet solidus from the bulk composition JSP 6-2 of ref. [6].

Diffusion modelling of Ti concentration profiles in quartz was performed using a finite-difference numerical integration of the non-steady-state diffusion equation for a spherical geometry[30]. In this model, a spherical quartz crystal with a 2-mm radius contains a discrete step in Ti concentration at the 1-mm radius (Fig. 3a). The model starts at initial temperature T_i and is allowed to cool at a constant rate to room temperature. Diffusivity was recalculated at each time step to accommodate changes during cooling.

Diffusion modelling is based on the assumption that blue-wavelength cathodoluminescence intensity (at about 452 nm) is directly proportional to the Ti concentration in the quartz crystal[34]. Cathodoluminescence maps contained full wavelength cathodoluminescence spectra for each map pixel which enabled selection of the specific wavelength at which Ti luminesces in quartz, thereby reducing the possibility that the cathodoluminescence intensities are conflated with cathodoluminescence signals at other wavelengths. This assumption, in addition to the normalization of the cathodoluminescence intensity profiles, allows changes in cathodoluminescence intensity to be used as a proxy for changes in Ti content of quartz. The higher resolution of cathodoluminescence images compared to Ti concentration profiles allows for fine-scale resolution of Ti gradients in quartz.

Results. Cathodoluminescence imaging reveals a range of zoning patterns in quartz, including nearly homogeneous single grains (Fig. 1a); gradational core–rim decreases in Ti (Fig. 1b); and growth (oscillatory) zoning (Fig. 1c,d). For example, individual quartz crystals from the Glen Aulin tonalite were relatively homogeneous in composition (although the average concentrations varied by 100 p.p.m. Ti between crystals), whereas quartz from the Cathedral Peak granodiorite exhibited growth zoning and dissolution–recrystallization textures similar to Ba-zoning in K-feldspar megacrysts from the same rock[35]. Regardless of internal zoning structure, quartz crystals from the TIS exhibit a rimward decrease in Ti concentration. Consistently, rim compositions of quartz crystals from all granitic rocks reached Ti contents between 20 and 40 p.p.m. Ti (see Source Data for Extended Data Fig. 3). The centres of quartz crystals (not necessarily the cores) ranged from about 50 p.p.m. to about 160 p.p.m. Ti. Secondary fluorescence of Ti in the quartz was observed in quartz crystals adjacent to titaniferous phases (for example titanite; Extended Data Fig. 2). No secondary fluorescence was observed in quartz crystals adjacent to Ti-poor minerals (for example plagioclase and K-feldspar).

Secondary fluorescence. Line transects towards the rims of quartz crystals adjacent to titaniferous phases (such as titanite) resulted in erroneously high Ti concentrations as a result of secondary fluorescence. Secondary fluorescence effects were observed at distances up to 100 μm from the rim of the quartz crystal (Extended Data Fig. 2), consistent with reports from other studies[18]. Importantly, cathodoluminescence zoning in the quartz crystal is relatively flat along the transect and does not show an increase in Ti in the quartz, proving that the increase in observed Ti is due to secondary fluorescence and not subsolidus kinetic diffusion of Ti into the quartz lattice.

Subsolidus re-equilibration and recrystallization. Subsolidus kinetic re-equilibration as a means to alter the trace element content of quartz can be discounted in these samples by looking at cathodoluminescence zoning in the crystals. For example, quartz from the Cathedral Peak granodiorite exhibits growth zoning in the blue wavelength consistent with growth of the crystal from a melt or fluid (Fig. 2c,d). The sharp boundaries between the growth zones in this and other crystals suggest no measurable kinetic re-equilibration, which would blend regions of high and low Ti in the quartz crystals resulting in 'smearing' of the blue cathodoluminescence blue signal. Cathodoluminescence images show that some quartz crystals (Extended Data Fig. 3) have undergone disruption to their initial Ti through shearing during magma emplacement, post-crystallization dynamic recrystallization or dissolution–reprecipitation. None of the quartz crystals that appear to have been altered by non-magmatic processes were included in the discussion of this paper.

Low crystallization temperatures are often interpreted as either post-magmatic recrystallization through hydrothermal alteration[12] or solid-state recrystallization of granites[36]. Post-magmatic hydrothermal alteration of the TIS can be discounted because the rocks investigated in this study contain none of the typical hallmarks of hydrothermal alteration of granites (for example quartz leaching, albitization and chloritization[37]).

Diffusion modelling of cathodoluminescence profiles shows that appreciable solid-state diffusive re-equilibration of Ti in quartz can be discounted because the sharp boundaries between growth zones documented here would have been blurred by diffusive re-equilibration (Fig. 3) or would exhibit solid-state metamorphic textures. The only sign of subsolidus recrystallization observed in quartz can be seen in the cathodoluminescence map of quartz from the Cathedral Peak granodiorite (Kcp, Fig. 2c), in which a portion of the rim of the quartz crystal contains a cathodoluminescence pattern akin to solid-state metamorphic crystallization of quartz[21]. This is an easily recognizable feature in cathodoluminescence maps and accounts for only a small fraction (<10%) of the total quartz population; therefore, solid-state recrystallization has affected a volumetrically insignificant amount of the TIS.

Choice of Ti-in-quartz thermobarometer calibration. Several iterations of the Ti-in-quartz thermobarometer have been developed, and the choice of thermobarometer can influence temperature and pressure estimates[18,20,38,39]. For this study, we use the original pressure-dependent calibration of Ti solubility in quartz[18]. Several criticisms have been aimed at the calibration in ref. [18], including poor matches to natural systems whose T and P were determined by independent means[40]; and lack of thermodynamic or other justification for the growth-rate dependence of Ti solubility[20]. In their response to ref. [40], Thomas et al.[18] note that the systems investigated by Wilson et al.[40] have poorly constrained $a_{TiO_2}^{melt-rutile}$ and probably disequilibrium growth of the Fe–Ti oxides, which had previously been demonstrated[41]. The growth-rate dependence on Ti uptake in quartz suggested by ref. [20] was proven inconsequential to the equilibrium state of the ref. [18] experiments in a set of reversal experiments and growth-entrapment modelling[42,43]. The calibration at low pressures by Ostapenko et al.[39] is in good agreement with the down-P extrapolation of ref. [18] but does not extrapolate well to high pressures. Additionally, as demonstrated in Fig. 3, Ti diffusion models agree with the ref. [18] calibration. Higher crystallization temperatures, as would be given for 20–40 p.p.m. Ti in quartz using the calibration by Huang and Audétat[20] (Extended Data Fig. 4) or by Wark and Watson[38], would cause diffusive relaxation of Ti-in-quartz concentration profiles that are not observed in the cathodoluminescence maps. For these reasons, we maintain that the ref. [18] calibration is the most robust calibration so far.

TiO_2 activity. An estimate of TiO_2 activity (hereafter designated $a_{TiO_2}^{melt-rutile}$) during quartz crystallization is necessary for the application of Ti-in-quartz thermobarometry in rutile-undersaturated systems. Unity TiO_2 activity of the melt ($a_{TiO_2}^{melt-rutile} = 1$) is defined relative to rutile saturation in the melt such that—at

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

rutile saturation—adding more TiO_2 to the system will result in the precipitation of an equimolar quantity of rutile and no increase in the TiO_2 content of the melt. In most silicic magmatic systems (including the TIS), a lack of rutile indicates that $a_{TiO_2}^{melt-rutile}$ is lower than 1. The presence of titanite throughout the TIS suggests $a_{TiO_2}^{melt-rutile} \approx 0.5–0.6$ (refs [44,45]). Others have argued for $a_{TiO_2}^{melt-rutile} \approx 0.75–0.8$ for the same rocks (for example see ref. [25]), which would lower the calculated Ti-in-quartz temperatures. If quartz begins crystallizing before titanite saturation in the magmas[46], it is reasonable to assume that interior portions of the quartz crystals could have formed at $a_{TiO_2}^{melt-rutile}$ values less than 0.5–0.6. For example, if, instead of $a_{TiO_2}^{melt-rutile} = 0.6$, the centre of the quartz crystal from Kcp (Fig. 1c) crystallized at $a_{TiO_2}^{melt-rutile} = 0.2$, as has been suggested for some titanite-free volcanic systems[41,47], the temperature estimate for about 60 p.p.m. Ti at 2 kbar pressure would increase from 572 °C to 687 °C. The value of $a_{TiO_2}^{melt-rutile} = 0.5–0.6$ used in this manuscript is a conservative value (at the low end of $a_{TiO_2}^{melt-rutile}$ estimates for titanite-bearing silicic rocks). Higher $a_{TiO_2}^{melt-rutile}$ estimates would further decrease the temperatures calculated for quartz rims. If TiO_2 activity in rutile-undersaturated melts is much lower than current estimates, solidus curves (for example, Fig. 2) could intersect the rim concentrations of quartz crystals. Low TiO_2 activity is possible if Ti in melt behaves in a non-Henrian manner.

Possible range of temperatures. Titanium concentrations of 40 p.p.m. at the rims of quartz crystals and an $a_{TiO_2}^{melt-rutile} = 0.5$ in the TIS (the combined rim Ti concentration and $a_{TiO_2}^{melt-rutile}$ estimate that would yield the highest temperature estimate) would need to crystallize at pressures of at least 6.3 kbar (approximately 19 km depth) for the wet solidus to agree with the observed Ti concentrations. Pressures this high are unlikely given that they are greater than the current pressure estimates, the highest of which is about (3.5 ± 0.5) kbar for Al-in-hornblende from the Half Dome granodiorite[14]. Alternatively, $a_{TiO_2}^{melt-rutile}$ would need to be near or below about 0.15. So low a value is unlikely given the ubiquity of titanite in the TIS, which implies relatively high $a_{TiO_2}^{melt-rutile}$ (refs [44,45]). We use pressure estimates of 1.6–2.4 kbar, based on Al-in-hornblende[14,31,48], and $a_{TiO_2}^{melt-rutile} = 0.5–0.6$ to obtain the range of temperature estimates given within the manuscript. Although the errors are relatively large on pressure estimates, measured temperatures would remain below the solidus unless the actual pressure of crystallization was 2–3 times higher than all previous geobarometric estimates.

Amount of low-temperature quartz. Most quartz in the Cathedral Peak granodiorite crystallized at low temperatures. Diffusion modelling of Ti concentration profiles in quartz from the Cathedral Peak granodiorite indicates that almost all the quartz from this unit crystallized at temperatures lower than the traditional solidus of about 700 °C (Extended Data Fig. 5). Furthermore, even at very rapid cooling rates (for example, 100 °C Myr^{-1}), the diffusion models cannot match observed concentration gradients if the initial temperature is at the wet solidus (Extended Data Fig. 6).

The diffusion modelling approach only works on crystals for which one can assume an initial step-function concentration gradient for Ti. This is often not the case. In the Cathedral Peak granodiorite, Ti concentrations reach as high as 60 p.p.m. Ti, and diffusion modelling demonstrates that this quartz crystallized at low temperatures. Using 60 p.p.m. as a conservative upper limit on low-T quartz crystallization, this value can then be used to assess the extent of low-T crystallization in quartz from samples where diffusion modelling is not possible. For example, using a spherical approximation for quartz growth in a crystal from the Half Dome granodiorite demonstrates that even with relatively large high-Ti cores, the main volume of quartz (about 90%) crystallized at low temperatures throughout the TIS (Extended Data Fig. 7).

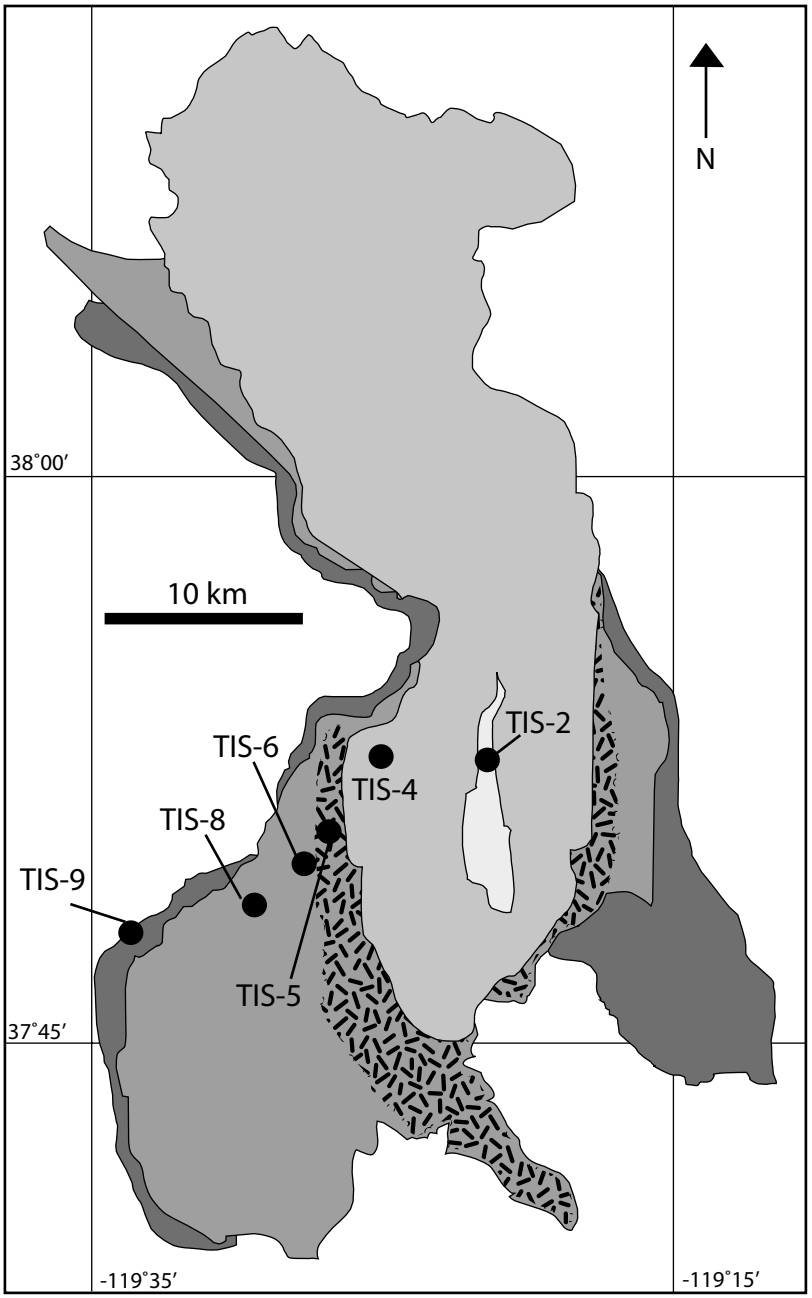
Other low-temperature indicators. In addition to quartz, the composition of the feldspars[12,35], oxygen isotope fractionation between quartz and zircon[19], and greenschist-facies mineralogy recorded within amphiboles[13] all indicate crystallization below the traditional wet solidus (Extended Data Fig. 8).

Code availability. Diffusion model code is available upon request from the corresponding author.

Data availability. All data used in this study are available in the Extended Data and Source Data files.

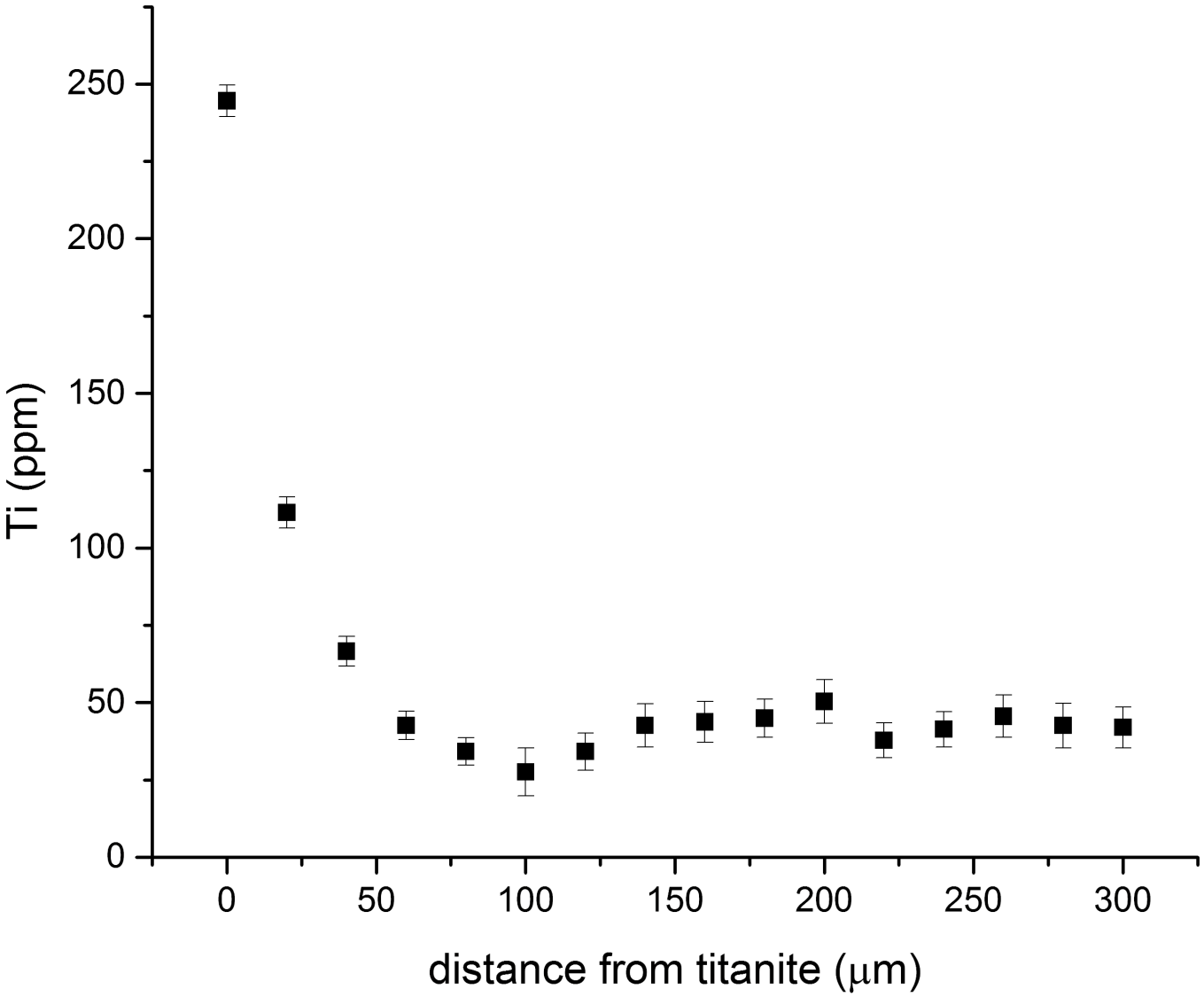
31. Ague, J. J. & Brimhall, G. H. Magmatic arc asymmetry and distribution of anomalous plutonic belts in the batholiths of California: effects of assimilation, crustal thickness, and depth of crystallization. *Geol. Soc. Am. Bull.* **100**, 912–927 (1988).
32. Memeti, V., Paterson, S., Matzel, J., Mundil, R. & Okaya, D. Magmatic lobes as 'snapshots' of magma chamber growth and evolution in large, composite batholiths: an example from the Tuolumne intrusion, Sierra Nevada, California. *Geol. Soc. Am. Bull.* **122**, 1912–1931 (2010).
33. Leeman, W. P. et al. A study of cathodoluminescence and trace element compositional zoning in natural quartz from volcanic rocks: mapping titanium content in quartz. *Microsc. Microanal.* **18**, 1322–1341 (2012).
34. Spear, F. S., Ashley, K. T., Webb, L. E. & Thomas, J. B. Ti diffusion in quartz inclusions: implications for metamorphic time scales. *Contrib. Mineral. Petrol.* **164**, 977–986 (2012).
35. Johnson, B. R. & Glazner, A. F. Formation of K-feldspar megacrysts in granodioritic plutons by thermal cycling and late-stage textural coarsening. *Contrib. Mineral. Petrol.* **159**, 599–619 (2010).
36. Anderson, J. L. Status of thermobarometry in granitic batholiths. *Earth Environ. Sci. Trans. R. Soc. Edinb.* **87**, 125–138 (1996).
37. Hecht, L., Thuro, K., Plinninger, R. & Cuney, M. Mineralogical and geochemical characteristics of hydrothermal alteration and episyenitization in the Königshain granites, northern Bohemian Massif, Germany. *Int. J. Earth Sci.* **88**, 236–252 (1999).
38. Wark, D. A. & Watson, E. B. TitaniQ: a titanium-in-quartz geothermometer. *Contrib. Mineral. Petrol.* **152**, 743–754 (2006).
39. Ostapenko, G. T., Tarashchan, A. N. & Mitsyuk, B. M. Rutile–quartz geothermobarometer. *Geochem. Int.* **45**, 506–508 (2007).
40. Wilson, C. J. N., Seward, T. M., Allan, A. S. R., Charlier, B. L. A. & Bello, L. A comment on: 'TitaniQ under pressure: the effect of pressure and temperature on the solubility of Ti in quartz', by Jay B. Thomas, E. Bruce Watson, Frank S. Spear, Philip T. Shemella, Saroj K. Nayak and Antonio Lanzirotti. *Contrib. Mineral. Petrol.* **164**, 359–368 (2012).
41. Ghiorso, M. S. & Gualda, G. A. R. A method for estimating the activity of titania in magmatic liquids from the compositions of coexisting rhombohedral and cubic iron–titanium oxides. *Contrib. Mineral. Petrol.* **165**, 73–81 (2013).
42. Lanzillo, N. A., Watson, E. B., Thomas, J. B., Nayak, S. K. & Curioni, A. Near-surface controls on the composition of growing crystals: Car–Parrinello molecular dynamics (CPMD) simulations of Ti energetics and diffusion in alpha quartz. *Geochim. Cosmochim. Acta* **131**, 33–46 (2014).
43. Thomas, J. B., Watson, E. B., Spear, F. S. & Wark, D. A. TitaniQ recrystallized: experimental confirmation of the original Ti-in-quartz calibrations. *Contrib. Mineral. Petrol.* **169**, 27 (2015).
44. Watson, E. B., Wark, D. A. & Thomas, J. B. Crystallization thermometers for zircon and rutile. *Contrib. Mineral. Petrol.* **151**, 413–433 (2006).
45. Hayden, L. A. & Watson, E. B. Rutile saturation in hydrous siliceous melts and its bearing on Ti-thermometry of quartz and zircon. *Earth Planet. Sci. Lett.* **258**, 561–568 (2007).
46. Glazner, A. F., Coleman, D. S. & Bartley, J. M. The tenuous connection between high-silica rhyolites and granodiorite plutons. *Geology* **36**, 183–186 (2008).
47. Thomas, J. B. & Watson, E. B. Application of the Ti-in-quartz thermobarometer to rutile-free systems. Reply to: a comment on: 'TitaniQ under pressure: the effect of pressure and temperature on the solubility of Ti in quartz' by Thomas et al. *Contrib. Mineral. Petrol.* **164**, 369–374 (2012).
48. Anderson, J. L. & Smith, D. R. The effects of temperature and fO_2 on the Al-in-hornblende barometer. *Am. Mineral.* **80**, 549–559 (1995).
49. Putirka, K. D. Thermometers and barometers for volcanic systems. *Rev. Mineral. Geochem.* **69**, 61–120 (2008).
50. Trail, D., Bindeman, I. N., Watson, E. B. & Schmitt, A. K. Experimental calibration of oxygen isotope fractionation between quartz and zircon. *Geochim. Cosmochim. Acta* **73**, 7110–7126 (2009).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



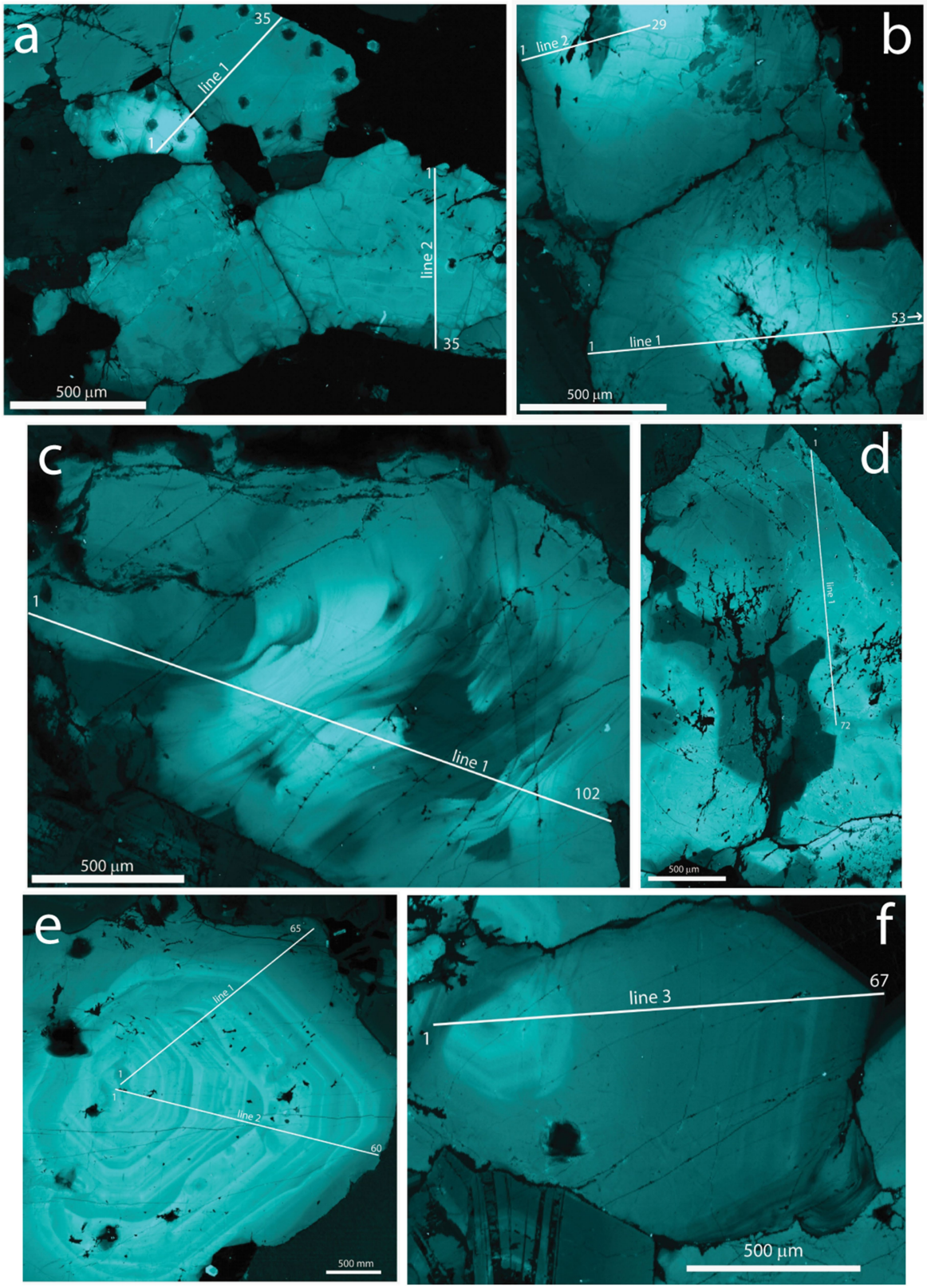
Extended Data Fig. 1 | Modified geological map of the TIS with sample locations. Geological map modified from ref. [10].

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



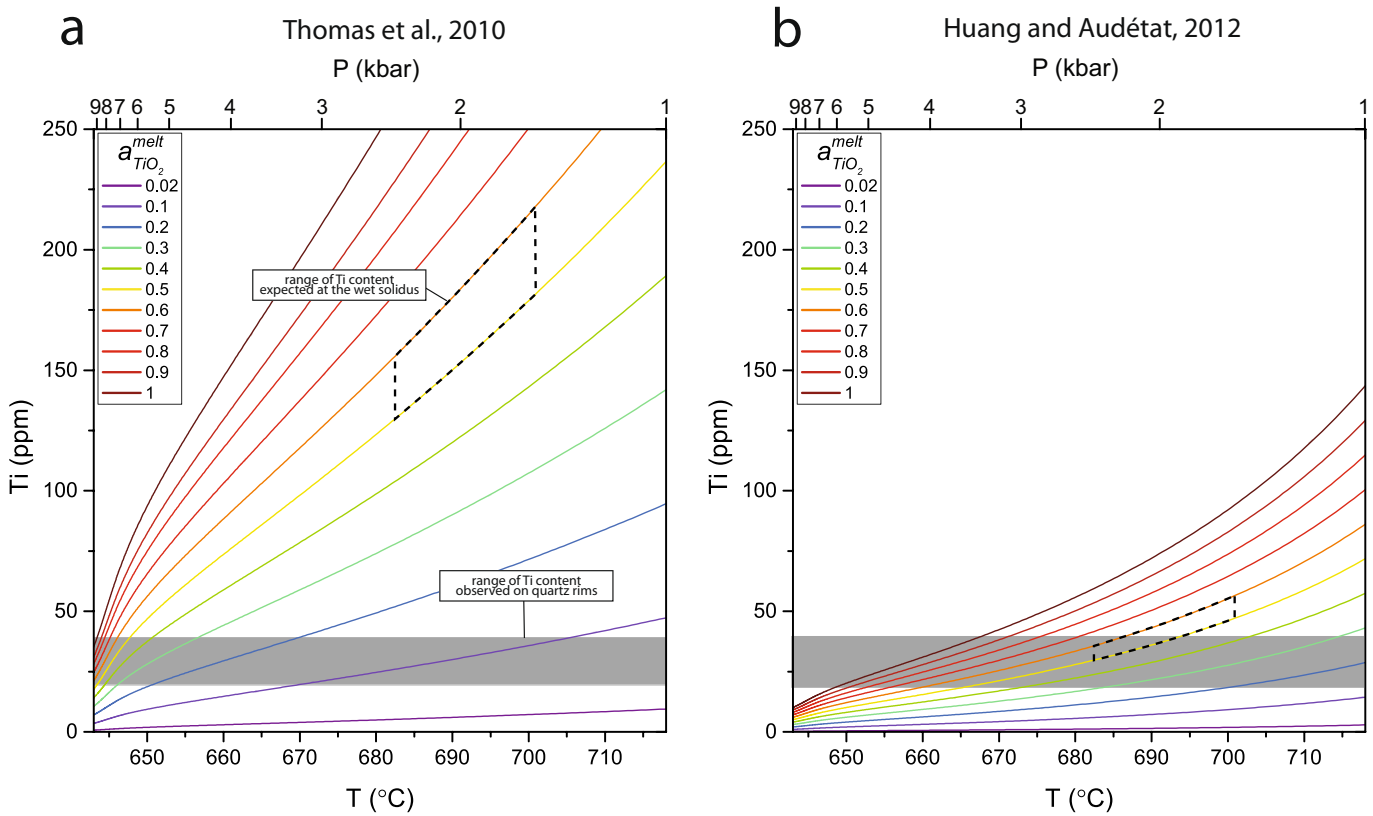
Extended Data Fig. 2 | Secondary fluorescence analyses of Ti concentrations in quartz as a function of distance from an adjacent titanite crystal. Analyses of the Ti of a quartz crystal from the Glen Aulin tonalite (Kga) show a rimward increase in Ti towards an adjacent titanite crystal. This increase in Ti content is due to a secondary fluorescence of Ti from the nearby titanite crystal. Error bars are 1σ.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



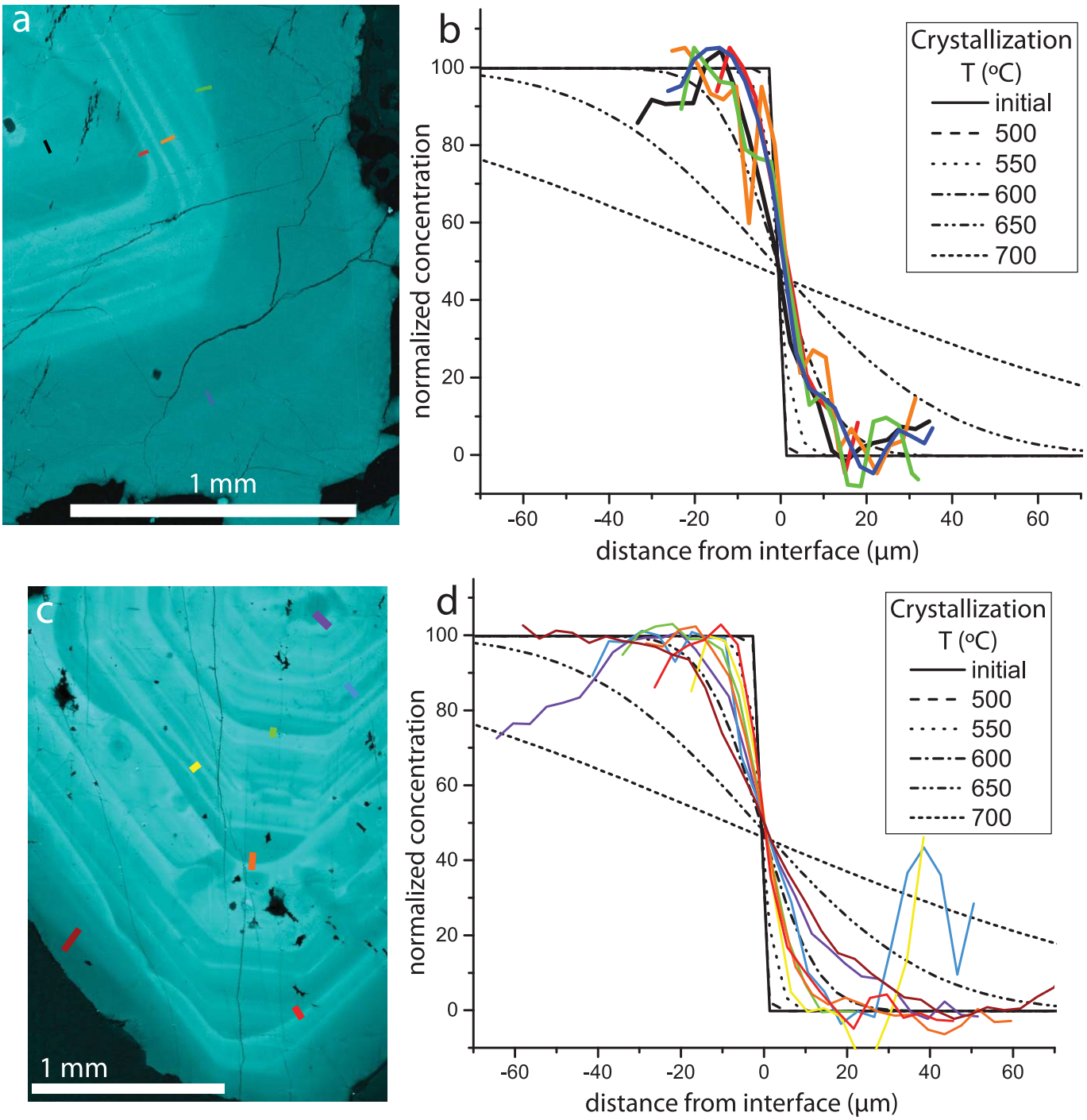
Extended Data Fig. 3 | Blue-filtered cathodoluminescence images of quartz from the TIS demonstrate both primary crystallization features and features that suggest alteration of the initial Ti content. a, Quartz from the Glen Aulin tonalite. **b**, Quartz from the equigranular Half Dome granodiorite has a centre-to-rim decrease in Ti content consistent with lower-temperature crystallization. **c**, **d**, In contrast, other quartz crystals from the equigranular Half Dome granodiorite exhibit complex cathodoluminescence features suggesting that the primary Ti content of the quartz has been altered by syn-magmatic deformation. **e**, **f**, Quartz from the Cathedral Peak granodiorite displays a series of primary magmatic growth features suggesting thermal cycling and dissolution-recrystallization of quartz. Lines are microprobe traverses.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



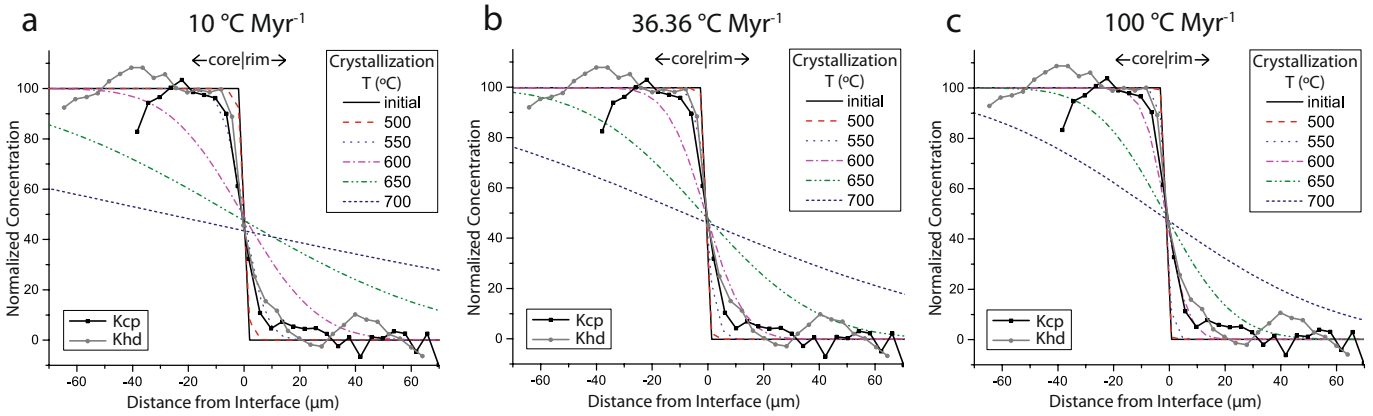
Extended Data Fig. 4 | Comparison of calibrations of the Ti-in-quartz thermobarometer. a, Calibration by Thomas et al.[18]; **b**, calibration by Huang and Audétat[20]. Although the Huang and Audétat[20] calibration could yield near-solidus temperatures for 20–40 p.p.m. Ti in quartz, these high temperatures cannot explain the sharp Ti concentration zones in quartz crystals, which could only be retained if the quartz crystallized at low temperatures (Fig. 3).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



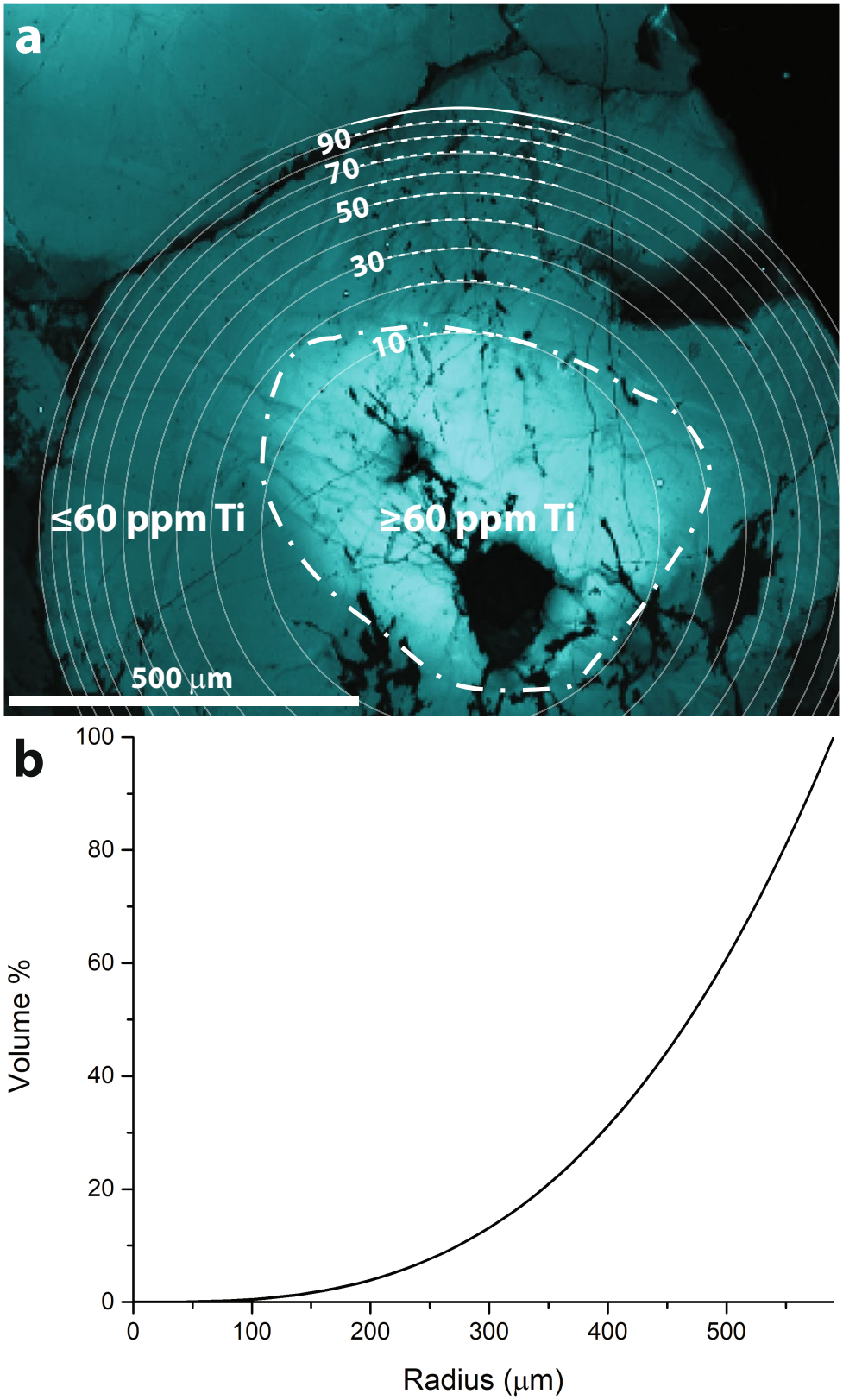
Extended Data Fig. 5 | Core–rim Ti concentration profiles in quartz compared with a constant-cooling rate diffusion model. Diffusion model from Fig. 3b. **a**, Cathodoluminescence (452 nm) map image of a quartz crystal from the Cathedral Peak granodiorite from approximately the core (upper left) to the rim (right and bottom). Coloured lines are profiles seen in **b**. **b**, Normalized concentration gradients from profiles across the grain combined with the diffusion model from Fig. 3b indicate that initial crystallization temperatures must have been below the traditional solidus at about 700 °C for almost the entirety of the observed quartz crystal. For simplicity, high concentrations are all plotted on the left, regardless of position within the crystal. At the scale of these profiles, this simplification does not noticeably change the modelled profiles (that is, it does not matter if the high concentration is coreward or rimward). A second example from the Cathedral Peak granodiorite also contains core (top) to rim (bottom) cathodoluminescence (**c**) and corresponding concentration gradient profiles that indicate low temperature crystallization for the entire crystal (**d**).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



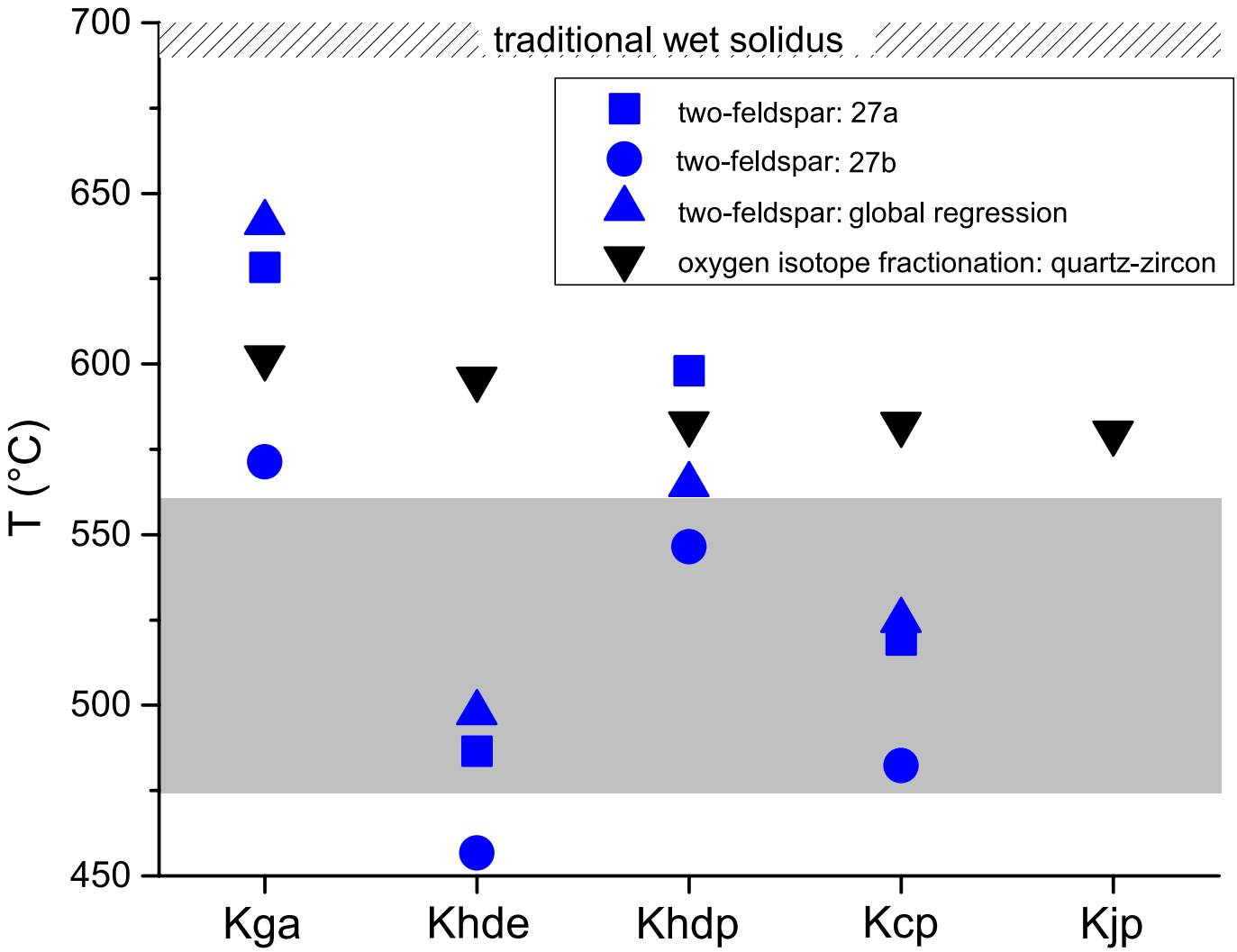
Extended Data Fig. 6 | The influence of cooling rate on the results of constant-cooling rate diffusion models. a–c, Diffusion modelling at 10 °C per million years (10 °C Myr^{-1}; **a**), 36.36 °C Myr^{-1} (**b**) and 100 °C Myr^{-1} (**c**). Faster cooling rates will lead to less diffusion. However, even at rapid cooling rates (for example, 100 °C Myr^{-1}), the modelled profiles do not match the observed Ti concentrations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 7 | Estimating the volume of low-temperature quartz crystallization when diffusion modelling is intractable. **a**, Cathodoluminescence (452 nm) image of a centre-cut quartz crystal from the Half Dome granodiorite. This crystal has internal Ti concentrations as high as 101 p.p.m. Concentric circles are contours showing the percentage of total volume crystallized at that radius, assuming spherical quartz growth. Using 60 p.p.m. Ti as a rough approximation for high-T quartz crystallization (denoted with a bold dot-dashed polygon), this volume approximation demonstrates that only around 10% of the quartz could have crystallized at solidus or supersolidus temperatures. **b**, Radius of crystal (approximately 590 μm) versus the total volume of quartz crystallized at a given radius for a spherical quartz crystal with $r = 590$ μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 8 | Low-temperature indicators from other minerals in the TIS. In addition to quartz, feldspar compositions and oxygen isotope fractionation between quartz and zircon results in subsolidus temperatures. Two-feldspar temperatures are calculated using three different calibrations (equations (27a), (27b) and a global regression from ref. [49]). Zircon–quartz oxygen isotope fractionation temperatures are calculated using averaged oxygen isotope compositions from quartz and zircon from individual units in the TIS[19] and an experimentally calibrated zircon–quartz fractionation relationship[50].

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Locations of samples analysed for this study

Sample Number	Unit	Map Symbol	Lat	Long
TIS-02	Johnson Granite Porphyry	Kjp	37°52'33.94"N	119°21'22.77"W
TIS-04	Cathedral Peak Granodiorite	Kcp	37°52'37.02" N	119°25'0.14"W
TIS-05	porphyritic Half Dome	Khd	37°50'39.94" N	119°26'49.6"W
TIS-06	aplite in Half Dome	-	37°49'33.96" N	119°28'14.53"W
TIS-08	Half Dome Granodiorite	Khd	37°48'44.59" N	119°29'23.03"W
TIS-09	Glen Aulin tonalite	Kga	37°48'04.24" N	119°33'43.23"W

Locations are reported using the WGS84 Coordinate System.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

https://doi.org/10.1038/s41586-018-0262-4

Hippocampal neurogenesis confers stress resilience by inhibiting the ventral dentate gyrus

Christoph Anacker[1]*, Victor M. Luna[1], Gregory S. Stevens[1], Amira Millette[1], Ryan Shores[1], Jessica C. Jimenez[1], Briana Chen[1] & René Hen[1,2,3]*

Adult neurogenesis in the dentate gyrus of the hippocampus is highly regulated by environmental influences, and functionally implicated in behavioural responses to stress and antidepressants[1–4]. However, how adult-born neurons regulate dentate gyrus information processing to protect from stress-induced anxiety-like behaviour is unknown. Here we show in mice that neurogenesis confers resilience to chronic stress by inhibiting the activity of mature granule cells in the ventral dentate gyrus (vDG), a subregion that is implicated in mood regulation. We found that chemogenetic inhibition of adult-born neurons in the vDG promotes susceptibility to social defeat stress, whereas increasing neurogenesis confers resilience to chronic stress. By using in vivo calcium imaging to record neuronal activity from large cell populations in the vDG, we show that increased neurogenesis results in a decrease in the activity of stress-responsive cells that are active preferentially during attacks or while mice explore anxiogenic environments. These effects on dentate gyrus activity are necessary and sufficient for stress resilience, as direct silencing of the vDG confers resilience whereas excitation promotes susceptibility. Our results suggest that the activity of the vDG may be a key factor in determining individual levels of vulnerability to stress and related psychiatric disorders.

The hippocampus is functionally heterogeneous along its dorsal–ventral axis[5]. Whereas the dorsal pole predominantly regulates cognition, the ventral pole has been implicated in stress responses and anxiety[6,7]. Within the dentate gyrus region of the hippocampus, young granule neurons continue to be generated in adulthood[6,8–10]. These adult-born neurons undergo a critical developmental period of heightened synaptic plasticity at 4–6 weeks post mitosis, during which they exert distinct contributions to behaviour[11–13]. Consistent with the role of the ventral hippocampus in mood regulation, adult-born neurons mediate some of the behavioural effects of antidepressants[2,14], and protect against stress-induced neuroendocrine and behavioural impairments[1,15,16].

To investigate whether the activity of adult-born granule cells (abGCs) in the vDG is required to protect from stress-induced anxiety-like behaviours, we generated a new loss-of-function model to chemogenetically silence abGCs in vivo. We expressed the inhibitory designer receptor exclusively activated by designer drugs (DREADD), hM_4D_i, in abGCs by crossing a transgenic mouse line with a tamoxifen-inducible CreERT2 recombinase, expressed under the control of the *Nestin* gene promoter[17], to mice expressing STOP-floxed hM_4D_i[18] (*Nestin-creERT2*$^{+/-}$;*loxP-stop-mCherry-loxP-hM*$_4$*D*$_i$, referred to as *i-hM*$_4$*D*$_i$ mice). In vitro electrophysiology confirmed that the DREADD

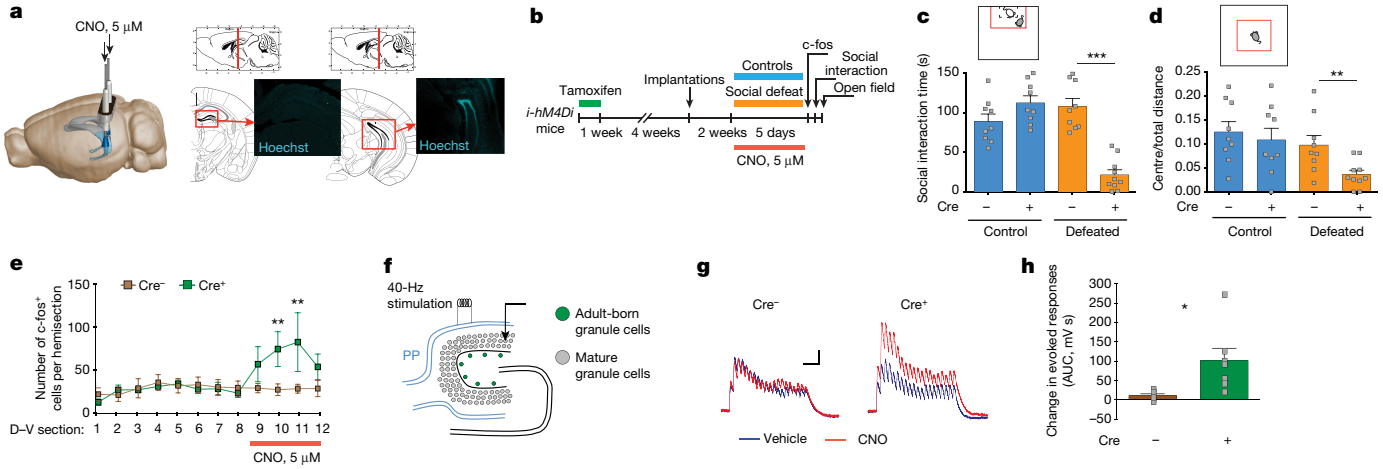


Fig. 1 | Silencing adult-born neurons promotes stress susceptibility and increases vDG excitability. a, Cannula placement and Hoechst33342 dye infusions into the vDG of *i-hM*$_4$*D*$_i$ mice. Image generated using the Allen Institute Brain Explorer 2 software (http://mouse.brain-map.org/static/brainexplorer). **b**, Experimental design for subthreshold defeat. **c**, Silencing abGCs in *i-hM*$_4$*D*$_i$ mice decreases social interaction time (interaction $F_{1,33} = 40.98$, $P < 0.001$; genotype $F_{1,33} = 13.46$, $P = 0.0009$; stress $F_{1,33} = 17.7$, $P = 0.0002$; post hoc test, defeat Cre^- versus Cre^+, $^{***}P < 0.001$; $n = 9, 9, 9, 10$). **d**, Silencing abGCs decreases open field centre exploration (interaction $F_{1,33} = 1.4$, $P = 0.25$; genotype $F_{1,33} = 4.2$, $P = 0.048$; stress $F_{1,33} = 6.9$, $P = 0.01$; planned comparison *t*-test, defeat Cre^- versus Cre^+, $^{**}P = 0.009$; $n = 9, 9, 9, 10$). **e**, Quantification of c-fos$^+$ cells (interaction $F_{11,176} = 2.1$, $P = 0.02$; genotype $F_{1,16} = 1.2$, $P = 0.28$; dorsal–ventral $F_{11,176} = 2.2$, $P = 0.01$; post hoc test, section 10: $^{**}P = 0.008$, section 11: $^{**}P = 0.002$, $n_{Cre^-} = 8$, $n_{Cre^+} = 10$). D–V, dorsal–ventral axis. **f**, Schematic of perforant path (PP) stimulation. **g**, Representative responses of mature granule cells (mGCs) to perforant path stimulation in vitro. Scale bars: 2 mV, 100 ms. **h**, Area under the curve (AUC) quantification of evoked responses (*t*-test, $^*P = 0.016$; $n = 7,7$). Error bars, ±s.e.m.

[1]Department of Psychiatry, Division of Systems Neuroscience, Columbia University and Research Foundation for Mental Hygiene, New York State Psychiatric Institute, New York, NY, USA. [2]Department of Neuroscience, Columbia University, New York, NY, USA. [3]Department of Pharmacology, Columbia University, New York, NY, USA. *e-mail: ca2635@columbia.edu; rh95@columbia.edu

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

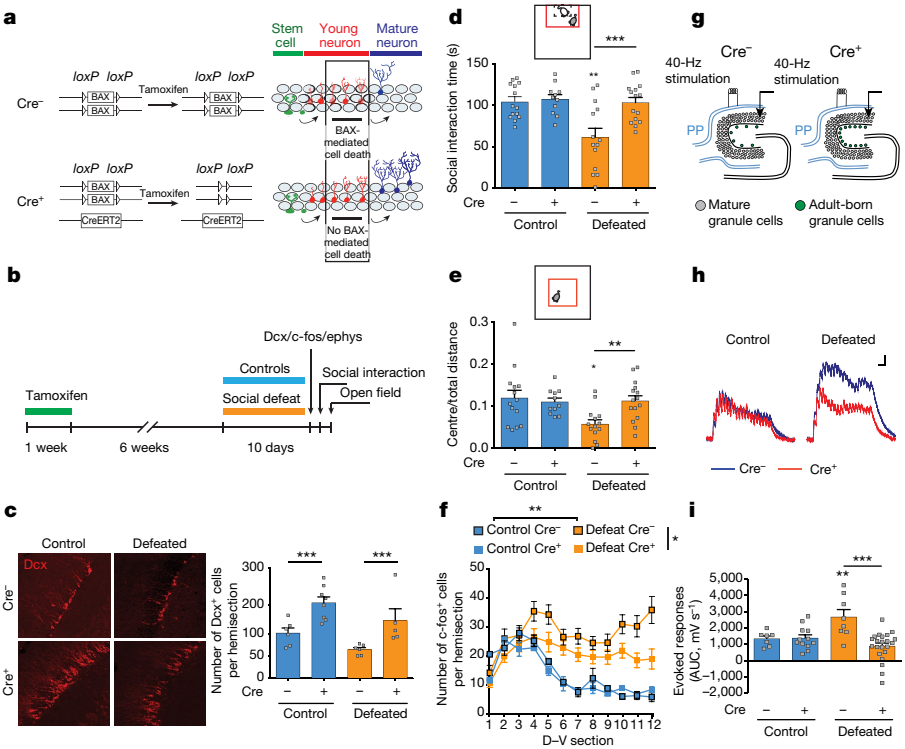


Fig. 2 | Increasing neurogenesis confers stress resilience and attenuates vDG excitability. a, Gain-of-function strategy to increase neurogenesis, resulting in *iBax* mice. **b**, Experimental design for chronic defeat. Ephys, electrophysiological recordings **c**, Quantification of Dcx^+ cells (interaction $F_{1,21} = 0.007$, $P = 0.9$; genotype $F_{1,21} = 20.1$, $^{***}P = 0.0002$; stress $F_{1,21} = 6.7$ $P = 0.02$; $n = 6, 8, 6, 5$). **d**, Social interaction time (interaction $F_{1,50} = 6.5$, $P = 0.014$; genotype $F_{1,50} = 9.0$, $P = 0.004$; stress $F_{1,50} = 9.5$, $^{**}P = 0.003$; post hoc test, defeat Cre^- versus Cre^+, $^{***}P = 0.002$; $n = 14, 11, 14, 15$). **e**, Open field centre exploration (interaction $F_{1,50} = 5.7$, $P = 0.021$; genotype $F_{1,50} = 2.9$, $P = 0.097$; stress $F_{1,50} = 4.9$, $^{*}P = 0.032$; post hoc test, defeat Cre^- versus Cre^+, $^{**}P = 0.004$; $n = 14, 11, 14, 15$). **f**, Quantification of c-fos^+ cells (interaction $F_{33,319} = 3.1$, $P < 0.001$; stress $F_{3,29} = 5.5$, $P = 0.004$; dorsal–ventral $F_{11,319} = 8.5$, $P < 0.0001$; post hoc test, Cre^- control versus defeat, $^{**}P = 0.007$; defeat Cre^- versus Cre^+, $^{*}P = 0.015$; $n_{control,Cre^-} = 3$, $n_{defeat,Cre^-} = 12$, $n_{control,Cre^+} = 3$, $n_{defeat,Cre^+} = 15$). Cre^+ *iBax* mice have less stress-induced c-fos^+ cells than Cre^- mice (interaction $F_{11,275} = 2.1$, $P = 0.02$; genotype $F_{1,25} = 5.8$, $P = 0.02$; dorsal–ventral $F_{11,275} = 10.7$, $P < 0.001$; $n_{defeat,Cre^-} = 12$, $n_{defeat,Cre^+} = 15$). **g**, Schematic of perforant path stimulation. **h**, Representative recordings from mGCs upon perforant path stimulation in vitro. Scale bars: 1 mV, 50 ms. **i**, AUC quantification of evoked responses (interaction $F_{1,45} = 11.05$, $P = 0.0018$; genotype $F_{1,45} = 10.1$, $P = 0.003$; stress $F_{1,45} = 2.3$, $P = 0.13$; post hoc tests, control Cre^- versus defeat Cre^-, $^{**}P < 0.005$; defeat Cre^- versus Cre^+, $^{***}P < 0.001$; $n = 7, 12, 8, 22$). Error bars, ±s.e.m.

receptor agonist clozapine-*N*-oxide (CNO) decreased abGC activity (Extended Data Fig. 1). We then silenced abGCs in vivo, using cannula-mediated delivery of CNO directly into the vDG every day during a five-day subthreshold social-defeat paradigm (Fig. 1a, b). This short paradigm was used because it is not sufficient to alter behaviour in Cre^- control mice[19]. However, silencing abGCs with CNO during this subthreshold social-defeat resulted in robust avoidance of a novel mouse in a social interaction test (Fig. 1c), and reduced centre exploration in the open field in defeated *i-hM4Di* mice that express hM_4D_i in young neurons (Cre^+) (Fig. 1d, Extended Data Fig. 2a–d).

To investigate how silencing abGCs affects the activity of the vDG in response to stress, we then examined expression of the immediate early gene *c-fos* (also known as *Fos*) as a proxy marker for neuronal activity. The number of c-fos^+ cells was selectively increased in the vDG upon silencing of abGCs during stress in vivo (Fig. 1e). Consistent with this effect, in vitro electrophysiological recordings from mature vDG granule cells showed increased evoked responses to perforant path stimulation after silencing abGCs with CNO (Fig. 1f–h). These results indicate that silencing young neurons increases the activity of mature granule cells in response to stress, and causes decreased social behaviour and increased anxiety-like behaviour.

Environmental influences that promote resilience to stress, such as environmental enrichment or voluntary exercise, have been shown to increase neurogenesis[20,21]. We therefore wanted to test whether increasing neurogenesis was sufficient to confer stress resilience. To do this, we used a gain-of-function model to increase neurogenesis by inducible deletion of the proapoptotic gene *Bax* from adult neural stem cells and their progeny (*Nestin-creERT2*$^{+/-}$;*Bax*$^{fl/fl}$, referred to as *iBax* mice[22]) (Fig. 2a). *iBax* mice were subjected to a chronic version of social defeat stress (10 days)[23] and subsequently tested in the social interaction test and in the open field (Fig. 2b). Deletion of *Bax* from adult neural stem cells (Cre^+) produced an approximately twofold increase in young Dcx^+ neurons (Fig. 2c) and an approximately sixfold increase in cell survival compared to control mice (Cre^-) (Extended Data Fig. 3a, b). In contrast to the five-day subthreshold defeat paradigm (Fig. 1b), chronic defeat for ten days (Fig. 2b) produces robust social avoidance and anxiety-like behaviour. Defeated Cre^- mice spent around 42% less time interacting with a novel mouse than did undefeated mice, while defeated *iBax* mice with increased neurogenesis (Cre^+) showed control levels of social interaction (Fig. 2d). Similar effects were observed in the open field, in which defeated Cre^- mice showed around 50% less exploration of the centre of the area than did undefeated mice, while defeated *iBax* mice exhibited control levels of centre exploration (Fig. 2e, Extended Data Fig. 2e–h). These results suggest that increased neurogenesis can confer resilience to chronic stress. To strengthen this conclusion, we showed that ablation of neurogenesis by focal X-ray irradiation of the vDG abolished these pro-resilience effects (Extended Data Fig. 3c–j), which confirms that abGCs located specifically in the vDG are responsible for the stress-resilience of *iBax* mice.

We next investigated the activity of the vDG in *iBax* mice. Chronic social defeat increased the number of c-fos^+ cells in the vDG (sections

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

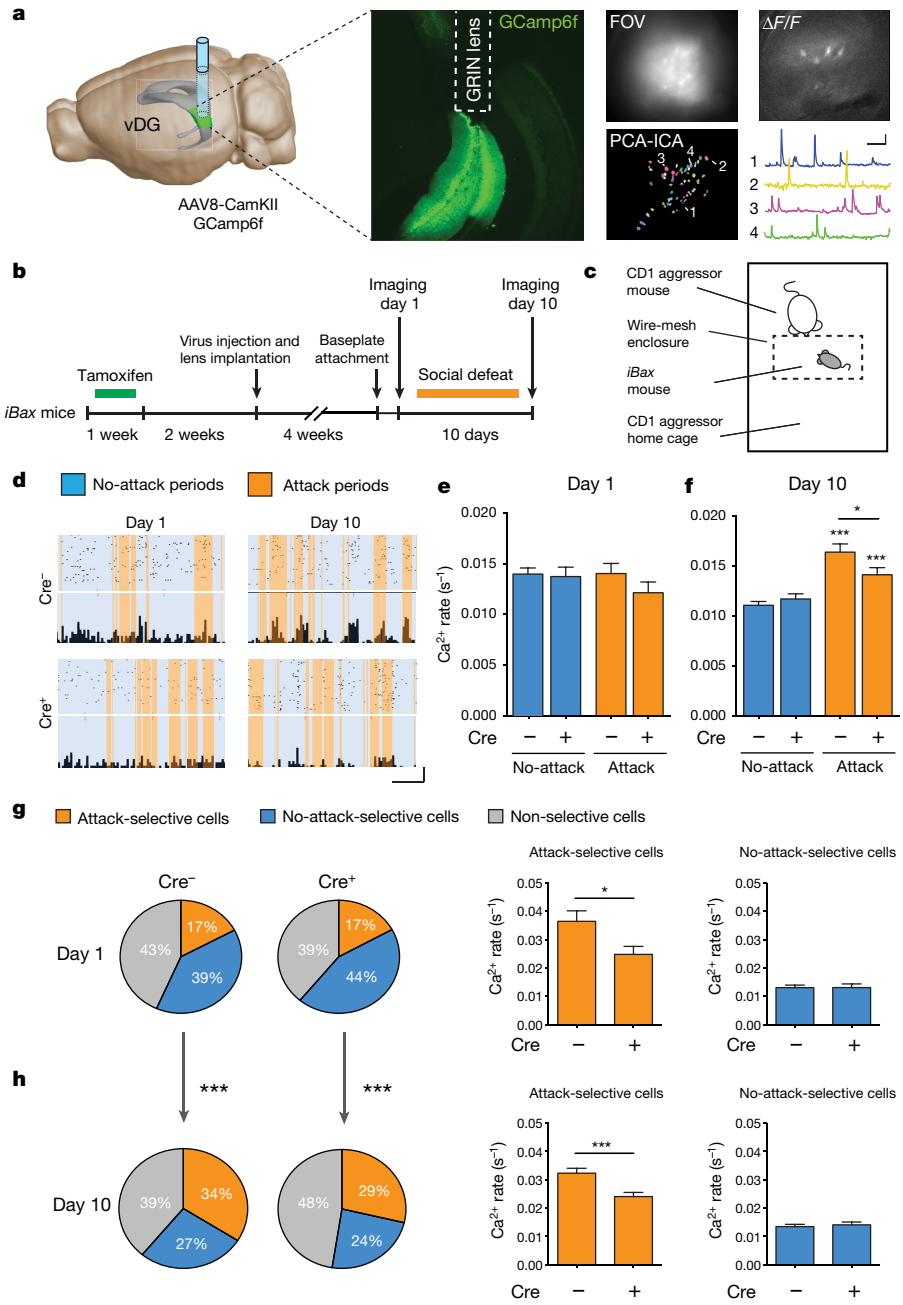


Fig. 3 | Neurogenesis inhibits vDG activity after chronic stress. a, Left, schematic illustrating gradient refractory index (GRIN) lens placement and GCamp6f expression. Image generated using the Allen Institute Brain Explorer 2 software. Right, representative images of field-of-view (FOV), $\Delta F/F$ fluorescence signals, and PCA-ICA (P/ICA, principal/independent component analysis); and Ca^{2+} transients. Scale bars: 50 s, 5 s.d. **b**, Experimental design. **c**, Mice were imaged in a protective wire mesh enclosure located inside the aggressor cage. **d**, Representative raster plots of Ca^{2+} transient events. Each line represents one cell. Bottom panels show cumulative Ca^{2+} transient rates. Scale bars: 50 s, 0.015 Ca^{2+} transient event rate. Representative 280-s periods from a 600-s recording are shown. **e**, Ca^{2+} transient rates on day 1 (no-attack: Mann–Whitney $U = 52{,}962$, $P = 0.16$; attack: $U = 55{,}260$, $P = 0.59$; $n_{Cre^-} = 460$, $n_{Cre^+} = 246$). **f**, Ca^{2+} transient rates are increased during attack periods on day 10 (Wilcoxon matched-pairs signed rank, Cre^- no-attack versus attack, $^{***}P_{Cre^-} < 0.0001$; $n_{Cre^-} = 620$; Cre^+ no-attack versus attack, $^{***}P_{Cre^+} = 0.0002$; $n_{Cre^+} = 592$). Cre^+ *iBax* mice show lower attack responses than Cre^- mice ($U = 169{,}211$, Cre^- attack versus Cre^+ attack $^{*}P = 0.017$). **g**, Cell selectivity on day 1. Ca^{2+} transient rates of attack-selective cells on day 1 ($U = 1{,}272$, $^{*}P = 0.027$; $n_{Cre^-} = 80$, $n_{Cre^+} = 42$). Ca^{2+} transient rates of no-attack-selective cells on day 1 ($U = 8{,}825$, $P = 0.19$; $n_{Cre^-} = 180$, $n_{Cre^+} = 108$). **h**, Cell selectivity analysis on day 10 showed more attack-selective cells compared to day 1 (chi-squared test: $\chi^2_{Cre-}(2) = 40.13$, Cre^- day 1 versus Cre^- day 10, $^{***}P_{Cre^-} < 0.0001$, $n_{Cre^-} = 620$; $\chi^2_{Cre+}(2) = 36.14$, Cre^+ day 1 versus Cre^+ day 10, $^{***}P_{Cre^+} < 0.0001$, $n_{Cre^+} = 592$). Ca^{2+} transient rates of attack-selective cells on day 10 ($U = 12{,}725$, $^{***}P < 0.0001$; $n_{Cre^-} = 211$, $n_{Cre^+} = 169$). Ca^{2+} transient rates of no-attack-selective cells on day 10 ($U = 11{,}544$, $P = 0.72$; $n_{Cre^-} = 169$, $n_{Cre^+} = 140$). Data distributions are shown in Extended Data Figs. 5 and 6. Error bars, ±s.e.m.

4–12) compared to undefeated controls, whereas this effect was attenuated in Cre^+ *iBax* mice (Fig. 2f). In vitro electrophysiological recordings in Cre^- mice showed increased evoked responses compared with undefeated controls. This effect was absent in *iBax* mice (Cre^+) (Fig. 2g–i). These findings show that increased neurogenesis reduces chronic-stress-induced increases in vDG activity, which may in turn protect against anxiety-like behaviours.

To determine how stress-related information is represented in the vDG and modulated by neurogenesis, we used in vivo calcium (Ca^{2+}) imaging of mature granule cells (Fig. 3a) while mice were attacked by an aggressor mouse through a protective wire-mesh enclosure (Fig. 3b, c). On the first day of defeat (day 1), no significant Ca^{2+} transient rate differences were observed between 'no-attack' and 'attack' periods (Fig. 3d, e). After chronic defeat (day 10), vDG granule cells showed increased Ca^{2+} transient rates during attack bouts compared to no-attack bouts in Cre^- mice. This effect was attenuated in *iBax* mice (Fig. 3d, f; Extended Data Fig. 4a–d).

To assess whether mature vDG granule cells display heterogeneous responses to attacks, we then used cell-selectivity analyses to identify a subset of cells (17%) in Cre^- mice and in *iBax* mice (Cre^+) that were active selectively during attack bouts on the first day of defeat (Fig. 3g). These attack-selective cells exhibited lower Ca^{2+} transient rates during attack periods in *iBax* mice than in Cre^- mice (Fig. 3g, orange bars). No rate differences were observed in no-attack-selective cells during periods of no attacks (Fig. 3g, blue bars), or in 'non-selective' cells (Extended Data Fig. 5a, b). After chronic defeat (day 10), the fraction of attack-selective cells was significantly increased compared to day 1 (Cre^-, 34%; Cre^+ (*iBax*), 29%) (Fig. 3h). Again, Ca^{2+} transient rates of attack-selective cells during attack periods were lower in Cre^+ *iBax* mice than in Cre^- mice (Fig. 3h, orange bars), whereas no rate differences were observed in no-attack-selective cells during periods of no attacks (Fig. 3h, blue bars) or in non-selective cells (Extended Data Fig. 5c). These results suggest that increased neurogenesis in *iBax* mice results in a decrease in the activity of attack-selective cells in the vDG. Similar results were obtained when considering the area under significant Ca^{2+} transients, amplitudes, or cumulative frequency distributions (Extended Data Figs. 5, 6). This neurogenesis-dependent reduction in vDG activity during stress extended to a reduction in vDG activity

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

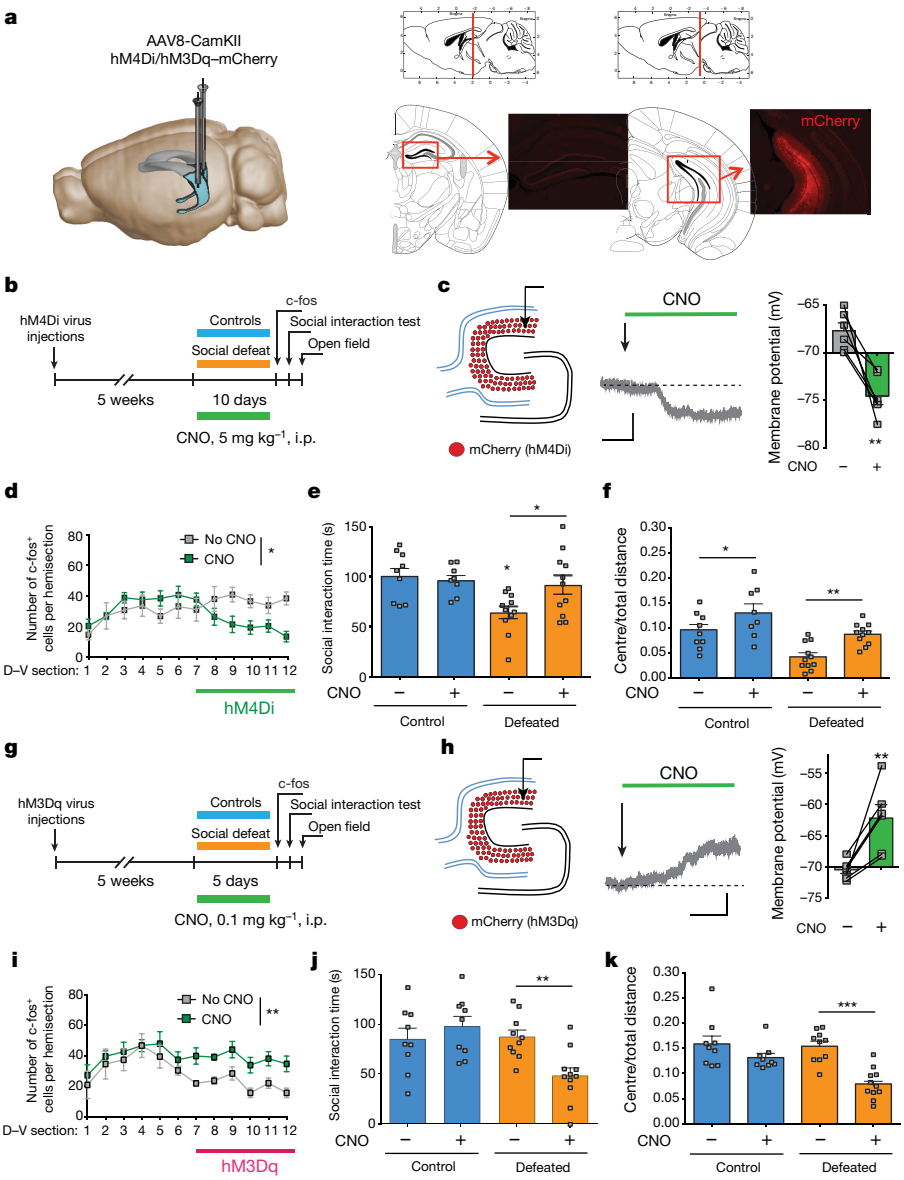


Fig. 4 | Direct modulation of vDG activity can regulate stress resilience. **a**, Virus injections and mCherry expression. Image generated using the Allen Institute Brain Explorer 2 software. **b**, Experimental design for vDG inhibition during chronic defeat (10 days). **c**, Resting membrane potential of hM_4D_i-infected cells, with and without CNO treatment (paired *t*-test, $^{**}P = 0.004$, $n = 6$). Scale bars: 5 mV, 2 min. **d**, Quantification of c-fos$^+$ cells after chronic defeat (sections 7–12: interaction $F_{5,40} = 3.5$, $P = 0.01$; CNO $F_{1,8} = 6.9$, $^*P = 0.03$; dorsal–ventral $F_{5,40} = 1.3$, $P = 0.29$; $n_{noCNO} = 4$, $n_{CNO} = 6$). **e**, Social interaction time (interaction $F_{1,35} = 4.2$, $P = 0.049$; CNO $F_{1,35} = 2.2$, $P = 0.15$; stress $F_{1,35} = 6.9$, $^*P = 0.013$; post hoc test, defeat vehicle versus CNO, $^*P = 0.01$; $n = 9, 8, 11, 11$). **f**, Open field centre exploration (interaction $F_{1,35} = 0.1$, $P = 0.73$; CNO $F_{1,35} = 14.4$, $P = 0.0006$; stress $F_{1,35} = 19.2$, $P = 0.0001$; post hoc test, control vehicle versus CNO, $^*P = 0.03$; defeat vehicle versus CNO, $^{**}P = 0.004$; $n = 9, 8, 11, 11$). **g**, Experimental design for vDG excitation during subthreshold defeat (5 days). **h**, Resting membrane potential of hM_3D_q-infected cells (paired *t*-test, $^{**}P = 0.009$, $n = 6$). Scale bars: 5 mV, 2 min. **i**, Quantification of c-fos$^+$ cells after subthreshold defeat (sections 7–12: interaction $F_{5,35} = 0.15$, $P = 0.98$; CNO $F_{1,7} = 12.3$, $^{**}P = 0.009$; dorsal–ventral $F_{5,35} = 4.8$, $P = 0.002$; $n_{noCNO} = 4$, $n_{CNO} = 5$). **j**, Social interaction time (interaction $F_{1,35} = 8.4$, $P = 0.007$; CNO $F_{1,35} = 2.1$, $P = 0.15$; stress $F_{1,35} = 6.9$, $P = 0.012$; post hoc test, defeat vehicle versus CNO, $^{**}P = 0.003$; $n = 9, 9, 10, 11$). **k**, Open field centre exploration (interaction $F_{1,35} = 4.7$, $P = 0.037$; CNO $F_{1,35} = 21.4$, $P < 0.0001$; stress $F_{1,35} = 6.6$, $P = 0.014$; post hoc test, defeat vehicle versus CNO, $^{***}P < 0.0001$; $n = 9, 9, 10, 11$). Error bars, ±s.e.m.

during stress-induced anxiety-like behaviour in the social interaction test and in the open field test (Extended Data Figs. 4e–o, 7, 8). Together, these results suggest that abGCs inhibit the activity of mature granule cells in the vDG, which in turn results in decreased anxiety-like behaviours after chronic stress.

Finally, to test whether vDG inhibition is sufficient to confer resilience, we directly silenced mature granule cells using viral-mediated expression of hM_4D_i (Fig. 4a, b). CNO hyperpolarized hM_4D_i-expressing granule cells in vitro (Fig. 4c) and reduced the chronic stress-induced increase in c-fos$^+$ cells in the vDG in vivo (Fig. 4d). Direct silencing of mature granule cells counteracted the reduction in social interaction time (Fig. 4e) and open field centre exploration (Fig. 4f) induced by chronic defeat, which suggests that vDG inhibition is sufficient to confer stress resilience. Conversely, direct excitation of mature vDG granule cells by activating virally expressed hM_3D_q receptors with CNO (Fig. 4g) depolarized hM_3D_q-expressing neurons in vitro (Fig. 4h) and increased the number of c-fos$^+$ cells in vivo after subthreshold defeat (Fig. 4i). This direct excitation resulted in reduced social interaction time (Fig. 4j) and open field centre exploration (Fig. 4k), which indicates that vDG

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

excitation is sufficient to promote stress susceptibility. In addition, the stress-susceptible phenotype of *i-hM$_4$D$_i$* mice was counteracted by simultaneously inhibiting mature granule cells, while the resilient phenotype of *iBax* mice was abolished by directly exciting mature granule cells (Extended Data Fig. 9). Overall, these data demonstrate that the behavioural effects of neurogenesis are mediated by a modulation of the activity of mature granule cells.

The determination of how neurogenesis regulates dentate gyrus information processing is crucial for our understanding of how adult-born neurons affect behaviour. Recent work has shown that adult-born neurons can inhibit mature granule cells via a mechanism that potentially implicates GABAergic hilar interneurons during hippocampus-dependent tasks[13,24–26]. Furthermore, the ventral CA1 contains cells that specifically encode anxiety-related information[27], and glutamatergic efferents from the ventral hippocampus can promote anxiety-like behaviour[27–29] (but see also Extended Data Fig. 10). Here we show that the vDG contains a population of cells that selectively respond to stressful stimuli, that this population of 'stress-responsive cells' is inhibited by adult-born neurons, and that this inhibition protects against the anxiogenic effects of a chronic social defeat paradigm.

Taken together, our results show that the regulation of vDG activity confers resilience to chronic stress, and suggest that strategies aimed at inhibiting the vDG may be protective against stress-related psychiatric disorders, such as anxiety and depression.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0262-4

Received: 23 August 2017; Accepted: 16 May 2018;
Published online 27 June 2018.

1. Snyder, J. S., Soumier, A., Brewer, M., Pickel, J. & Cameron, H. A. Adult hippocampal neurogenesis buffers stress responses and depressive behaviour. *Nature* **476**, 458–461 (2011).
2. Santarelli, L. et al. Requirement of hippocampal neurogenesis for the behavioral effects of antidepressants. *Science* **301**, 805–809 (2003).
3. Boldrini, M. et al. Antidepressants increase neural progenitor cells in the human hippocampus. *Neuropsychopharmacology* **34**, 2376–2389 (2009).
4. Lucassen, P. J., Stumpel, M. W., Wang, Q. & Aronica, E. Decreased numbers of progenitor cells but no response to antidepressant drugs in the hippocampus of elderly depressed patients. *Neuropharmacology* **58**, 940–949 (2010).
5. Strange, B. A., Witter, M. P., Lein, E. S. & Moser, E. I. Functional organization of the hippocampal longitudinal axis. *Nat. Rev. Neurosci.* **15**, 655–669 (2014).
6. Anacker, C. & Hen, R. Adult hippocampal neurogenesis and cognitive flexibility — linking memory and mood. *Nat. Rev. Neurosci.* **18**, 335–346 (2017).
7. Kheirbek, M. A. et al. Differential control of learning and anxiety along the dorsoventral axis of the dentate gyrus. *Neuron* **77**, 955–968 (2013).
8. Spalding, K. L. et al. Dynamics of hippocampal neurogenesis in adult humans. *Cell* **153**, 1219–1227 (2013).
9. Boldrini, M. et al. Human hippocampal neurogenesis persists throughout aging. *Cell Stem Cell* **22**, 589–599.e5 (2018).
10. Sorrells, S. F. et al. Human hippocampal neurogenesis drops sharply in children to undetectable levels in adults. *Nature* **555**, 377–381 (2018).
11. Danielson, N. B. et al. Distinct contribution of adult-born hippocampal granule cells to context encoding. *Neuron* **90**, 101–112 (2016).
12. Denny, C. A., Burghardt, N. S., Schachter, D. M., Hen, R. & Drew, M. R. 4- to 6-week-old adult-born hippocampal neurons influence novelty-evoked exploration and contextual fear conditioning. *Hippocampus* **22**, 1188–1201 (2012).
13. Temprana, S. G. et al. Delayed coupling to feedback inhibition during a critical period for the integration of adult-born granule cells. *Neuron* **85**, 116–130 (2015).
14. David, D. J. et al. Neurogenesis-dependent and -independent effects of fluoxetine in an animal model of anxiety/depression. *Neuron* **62**, 479–493 (2009).
15. Hill, A. S., Sahay, A. & Hen, R. Increasing adult hippocampal neurogenesis is sufficient to reduce anxiety and depression-like behaviors. *Neuropsychopharmacology* **40**, 2368–2378 (2015).
16. Culig, L. et al. Increasing adult hippocampal neurogenesis in mice after exposure to unpredictable chronic mild stress may counteract some of the effects of stress. *Neuropharmacology* **126**, 179–189 (2017).
17. Dranovsky, A. et al. Experience dictates stem cell fate in the adult hippocampus. *Neuron* **70**, 908–923 (2011).
18. Ray, R. S. et al. Impaired respiratory and body temperature control upon acute serotonergic neuron inhibition. *Science* **333**, 637–642 (2011).
19. Wook Koo, J. et al. Essential role of mesolimbic brain-derived neurotrophic factor in chronic social stress-induced depressive behaviors. *Biol. Psychiatry* **80**, 469–478 (2016).
20. Kempermann, G., Kuhn, H. G. & Gage, F. H. More hippocampal neurons in adult mice living in an enriched environment. *Nature* **386**, 493–495 (1997).
21. van Praag, H., Kempermann, G. & Gage, F. H. Running increases cell proliferation and neurogenesis in the adult mouse dentate gyrus. *Nat. Neurosci.* **2**, 266–270 (1999).
22. Sahay, A. et al. Increasing adult hippocampal neurogenesis is sufficient to improve pattern separation. *Nature* **472**, 466–470 (2011).
23. Krishnan, V. et al. Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions. *Cell* **131**, 391–404 (2007).
24. Burghardt, N. S., Park, E. H., Hen, R. & Fenton, A. A. Adult-born hippocampal neurons promote cognitive flexibility in mice. *Hippocampus* **22**, 1795–1808 (2012).
25. Ikrar, T. et al. Adult neurogenesis modifies excitability of the dentate gyrus. *Front. Neural Circuits* **7**, 204 (2013).
26. Drew, L. J. et al. Activation of local inhibitory circuits in the dentate gyrus by adult-born neurons. *Hippocampus* **26**, 763–778 (2016).
27. Jimenez, J. C. et al. Anxiety cells in a hippocampal-hypothalamic circuit. *Neuron* **97**, 670–683 (2018).
28. Padilla-Coreano, N. et al. Direct ventral hippocampal-prefrontal input is required for anxiety-related neural activity and behavior. *Neuron* **89**, 857–866 (2016).
29. Bagot, R. C. et al. Ventral hippocampal afferents to the nucleus accumbens regulate susceptibility to depression. *Nat. Commun.* **6**, 7062 (2015).

Acknowledgements This work was funded by the National Institutes of Health (K99 MH108719 to C.A; K01 AG054765 to V.M.L.; R37 MH068542, R01 MH083862, R01 NS081203 to R.H.), the Hope for Depression Research Foundation (HDRF RGA-13-003 to R.H.), NYSTEM (C029157 to R.H.) and the German Research Foundation (AN1006/1-1 to C.A.). G.S.S. and B.C. were supported by the Doctoral Program in Neurobiology and Behavior at Columbia University. J.C.J. is a Howard Hughes Medical Institute Gilliam Fellow and Columbia University MSTP student. We thank S. Dymecki for providing us with the Cre-responsive *loxP-stop-mCherry-loxP-hM$_4$D$_i$* mouse line.

Reviewer information *Nature* thanks P. Frankland and J. Snyder for their contribution to the peer review of this work.

Author contributions C.A. and R.H. designed the research and wrote the paper. C.A., A.M., B.C. and R.S. conducted behavioural experiments, surgeries, immunohistochemistry and cell counting. V.M.L. conducted in vitro electrophysiological experiments. C.A., G.S.S., J.C.J. and B.C. analysed in vivo imaging data.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0262-4.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0262-4.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to C.A. or R.H.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Mice. Procedures were conducted in accordance with the US National Institutes of Health Guide for the Care and Use of Laboratory Animals and New York State Psychiatric Institute Institutional Animal Care and Use Committee at Columbia University and the Research Foundation for Mental Hygiene.

Husbandry. Mice were housed in groups of 3–5 per cage with free access to food and water on a 12:12-h light/dark cycle. All behavioural testing was conducted during the light period.

Experimental mice. i-hM$_4$D$_i$ and *iBax* mice were bred in house. C57BL/6J wild-type mice were used for AAV8-CamKII-hM$_4$D$_i$–mCherry and AAV8-CamKII-hM$_3$D$_q$–mCherry injections and purchased from Jackson Laboratories. At least three separate cohorts of male mice were run for all experiments, each including Cre$^-$ and Cre$^+$ littermates from heterozygote breedings.

Aggressor mice. Single-housed CD1 retired breeder mice (Charles River, three months of age) were used as aggressors in social defeat experiments.

Data reporting. No statistical methods were used to predetermine sample size. Mice were randomly allocated to each experimental group by sampling across different litters to account for inter-litter effects in all experiments. Cagemates were randomly allocated to each experimental group by sampling across different cages. Three separate subsequent cohorts were run for all wild-type mice. Within these three separate cohorts, hM$_4$D$_i$ and hM$_3$D$_q$ experiments were run in parallel with mice randomly allocated to each experiment and each experimental group within each experiment. Each cohort consisted of 19–37 mice equally distributed across the hM$_4$D$_i$ and the hM$_3$D$_q$ experiment. Investigators quantifying Dcx and c-fos staining were blind to group. Investigators conducting behavioural experiments or manually scoring attack bouts were blind to genotype.

Viral constructs. AAV8-CamKII-GCamp6f was packaged and supplied by the Penn Vector Core at a titre of 4.3×10^{33} viral genomes (vg) per ml. Virus was diluted to a titre of 1.5×10^{13} vg ml^{-1} for injections. AAV8-CamKII-hM$_4$D$_i$–mCherry and AAV8-CamKII-hM$_3$D$_q$–mCherry were purchased from Addgene at a titre of $3–4 \times 10^{12}$ vg ml^{-1}.

Drugs. *Tamoxifen administration.* Eight-week-old control, *i-hM$_4$D$_i$* and *iBax* mice were administered tamoxifen (Sigma) pretreatment as indicated in the timelines shown in the figures. Tamoxifen was dissolved in 100% ethanol, suspended in corn oil and administered daily by intraperitoneal (i.p.) injection at 100 mg kg^{-1} for five consecutive days.

Clozapine-N-oxide (CNO) administration. For in vivo silencing of abGCs in *i-hM$_4$D$_i$* mice, CNO was dissolved in 100% dimethyl sulfoxide (DMSO) and diluted with 0.9% saline to a final concentration of 5 μM CNO and 0.001% DMSO. A volume of 400 nl CNO was infused bilaterally into the vDG every day 20 min before each social defeat session. To silence mGCs in the vDG, CNO (5 mg kg^{-1}) was injected intraperitoneally every day 30 min before each social defeat session into mice expressing AAV8-CamKII-hM$_4$D$_i$–mCherry. To excite mGCs, CNO (0.1 mg kg^{-1}) was administered by i.p. injection every day 30 min before each social defeat session into mice expressing AAV8-CamKII-hM$_3$D$_q$–mCherry in the vDG (see section entitled 'Viral injections').

Chronic social defeat stress. Chronic social defeat stress experiments were performed as previously described[30,31].

Aggression screening. CD1 retired breeders were screened for aggression before the start of the social defeat experiments. During the three-day screening procedure, a novel C57BL/6J mouse was introduced into the home cage of the CD1 mouse every day and the latency to attack during a 3-min testing session was recorded. CD1 mice that attacked on at least the two last screening days in less than 60 s were used as aggressors.

Subthreshold social defeat stress. For subthreshold social defeat stress, experimental mice were physically defeated by a new CD1 aggressor mouse for 5 min every day for 5 days. After every daily defeat session, experimental mice were housed across a perforated Plexiglass divider in the same cage with the aggressor for 24 h to allow continuous visual, olfactory and auditory contact, but no further physical defeat. Control mice were housed two per cage across a divider and paired with a novel control mouse every day. After 5 days, all mice were single-housed.

Chronic social defeat stress. For chronic exposure to social defeat stress, experimental mice were physically defeated as described above but for a total duration of 10 days. All mice were single-housed after the 10th day of defeat.

Behavioural testing. The social interaction test was conducted 24 h after the end of the last social defeat exposure on day 11 (for chronic defeat) or on day 6 (for subchronic defeat). The open field test was then conducted on day 12 (for chronic defeat) or on day 7 (for subchronic defeat). All behavioural testing was performed between 10:00 and 14:00.

Social interaction test. During this two-trial test, experimental mice were first placed into an arena (40 cm × 40 cm × 40 cm) for 150 s (trial 1). The arena contained an empty wire mesh enclosure (10 cm × 7 cm × 7 cm) alongside the middle of one of the arena walls. After trial 1, a novel CD1 mouse was placed into the wire-mesh enclosure and the experimental mouse was re-introduced into the arena for another 150 s (trial 2). Ethovision XT software (Noldus) was used to analyse the time that experimental mice spent in a social interaction zone that was defined to be 24 cm × 14 cm around the wire mesh enclosure with the new CD1 mouse during trial 2. In addition, the time mice spent in the corners (9 cm × 9 cm) opposite the social interaction zone was also recorded.

Open field test. Mice were placed in an open field arena (40 cm × 40 cm × 40 cm, Kinder Scientific) at 650 lux for 10 min. Behaviour was recorded with MotorMonitor software (Kinder Scientific) and the time mice spent in the centre of the open field arena (20 cm × 20 cm × 20 cm), as well as total distance travelled, were analysed.

Attack behaviour. Aggressive behaviour of CD1 mice towards experimental *iBax* mice during in vivo Ca^{2+} imaging was manually scored using Observer XT software (Noldus) by an experimenter blind to the experimental condition. Attacks were defined as biting the wire mesh enclosure containing the experimental mouse, and paw reaching into the enclosure.

Stereotactic surgeries. For all surgical procedures, mice were anaesthetized with 1.5% isoflurane at an oxygen flow rate of 1 l min^{-1} and positioned in a stereotaxic frame (David Kopf Instruments) on a T/pump warm water re-circulator (Stryker). Eyes were lubricated with opthalamic ointment. Fur was shaved and the incision site sterilized with betadine/ethanol swabs before surgery. Subcutaneous carprofen (5 mg kg^{-1}) was provided peri-operatively and for three days postoperatively for analgesia.

Bilateral cannula implantations. Bilateral guide cannulas for intra-hippocampal infusions of CNO were custom-made (Plastics One) to target the ventral dentate gyrus (vDG). The centre-to-centre spacing of the guide cannulas was 5.6 mm to fit a medial-lateral (ML) distance of 2.8 mm from the midline. Guide cannulas were cut at a length of 2.2 mm. Infusion cannulas were cut to fit the 2.2-mm guide cannulas with a 1-mm projection to reach a dorsal-ventral (DV) coordinate of 3.2 mm (from brain). Craniotomies were opened bilaterally with a dental drill at anterior-posterior (AP) −3.6 mm, and medial-lateral (ML) ±2.8 mm from the bregma line and midline, respectively. Cannulas were placed at an AP distance of −3.6 mm from bregma and slowly lowered into the brain with the infusion cannula attached to reach DV −3.2 mm. Two fully threaded 1/16″ microscrews (Antrin Miniature Specialities, Inc.) were implanted into the skull anterior to the cannula and the cannula was affixed to the skull and screws using Loctite 454 (Henkel). Infusion cannulas were removed and replaced by dummy cannulas of 2.2-mm length to fit the guide cannula without a projection into the vDG. Dummy cannulas stayed in place to protect guide cannulas from clogging. Mice were allowed to recover from surgery for two weeks before social defeat. Cannula placements into the vDG were confirmed using infusions of Hoechst33342 dye (1 μg ml^{-1}).

Viral injections. For AAV8-CamKII-hM$_4$D$_i$–mCherry and AAV8-CamKII-hM$_3$D$_q$–mCherry virus injections into the vDG, 400 nl of virus was bilaterally injected with a 10 μl NanoFil syringe and a 33g beveled needle (World Precision Instruments) at a constant speed of 100 nl min^{-1}. Craniotomies were performed as described above, and virus was injected bilaterally at −3.6 mm AP, ±2.8 mm ML and −3.2 mm DV. After each injection, the needle was left at the injection site inside the brain for an additional 5 min to aid diffusion from the needle tip and to prevent backflow. The needle was then slowly retracted and the scalp incision closed with Vetbond 3M. Each mouse was monitored and received carprofen (subcutaneous, 5 mg kg^{-1}) for pain management for three days after surgery. Mice were housed for five weeks postoperatively before the start of social defeat to allow for recovery from surgery and sufficient viral expression.

GRIN lens implantations. For in vivo Ca^{2+} imaging, mice underwent a single surgery in which 400 nl of AAV8-CamkII-GCamp6f virus was injected unilaterally with a 10-μl NanoFil syringe and a 33-g beveled needle (World Precision Instruments) at a constant speed of 100 nl min^{-1} before implanting a GRIN lens over the injection site. GRIN lenses were 6.1-mm long with a 0.5-mm diameter for vDG imaging (Inscopix). GRIN lens implantation was performed as previously described[27,32]. In brief, a craniotomy centred at the lens implantation site was made and dura was carefully removed from the brain surface and cleaned with a stream of sterile saline and absorptive spears (Fine Science Tools). No brain tissue was aspirated before lowering the GRIN lens. Three 1/16″ microscrews (Antrin Miniature Specialities, Inc.) were inserted in evenly spaced locations around the implantation site. The lens was lowered in 0.1-mm DV steps and then fixed to the skull with dental cement (Dentsply). Viral injections coordinates were −3.6 mm AP, ±2.8 mm ML, −3.2 mm DV (from brain). Lens placement coordinates were −3.6 mm AP, ± 2.8 mm ML, −2.9 mm DV (from brain). At the completion of surgery, the lens was protected with liquid mould rubber (Smooth On), and imaging experiments commenced four weeks later.

Focal X-ray irradiation of the ventral dentate gyrus. Irradiation was performed in a manner similar to that previously described[33]. In brief, six-week-old *iBax* mice were anaesthetized with 6 mg kg^{-1} sodium pentobarbital (Nembutal, Oak Pharmaceuticals), placed in a stereotaxic frame and rotated 90° clockwise along the coronal plane. A lead shield containing a 3.5 × 3.5 mm^2 sliding window was

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

positioned at 1 mm below to 2.5 mm above the interaural line to expose the ventral hippocampus to X-rays using an X-RAD 320 X-ray system (PXI) operated at 300 kV and 12 mA using a 2-mm Al filter. Three 2.5-Gy doses were delivered with a two-day inter-dose interval for a total of 7.5 Gy over a one-week period. Mice were allowed to recover for one week after irradiation and were then injected with tamoxifen as described above to induce Cre recombinase-mediated excision of the *BAX* gene in neural stem cells. Sham mice were anaesthetized but not irradiated. Chronic social defeat stress was carried out six weeks after tamoxifen injection. Two successive cohorts were run.

In vivo Ca^{2+} imaging. Four weeks after GRIN lens implantations, mice were checked for GCamp expression with a miniaturized microscope (Inscopix) as previously described[27,32]. Mice were briefly anaesthetized with 1.5% isoflurane at 1 l min^{-1} oxygen flow, and head-fixed into a stereotaxic frame. The protective rubber mould was removed from the lens and a magnetic baseplate was attached to the miniature microscope and lowered over the implanted GRIN lens to assess the FOV for GCamp6f-expressing neurons. If GCamp6f-expressing neurons were visible, the baseplate was affixed in place onto the headcap with dental cement. Ca^{2+} imaging sessions in freely moving mice were commenced one day after baseplating. For Ca^{2+} imaging on social defeat day 1 and day 10, experimental mice were physically defeated for 5 min by a CD1 aggressor mouse without the miniature microscope attached to the baseplate. After defeat, mice were briefly anaesthetized (<5 min) in order to attach the miniature microscope to the baseplate for imaging. Mice were allowed to recover from anaesthesia for 30 min and placed in a protective wire-mesh enclosure (13 cm × 7 cm × 9 cm) in the home cage of the aggressor. Mice were adapted to the aggressor cage and the imaging setup for 10 min without the CD1 aggressor present in the cage. The CD1 aggressor was then introduced into the cage for 10 min. After 10 days of social defeat, mice were imaged in the social interaction test (day 11) and in the open field test (day 12). Ca^{2+} videos were recorded with nVista acquisition software (Inscopix), and triggered with a TTL pulse from EthoVision XT10 and Noldus IO box system to allow for simultaneous acquisition of synchronized Ca^{2+} and behavioural videos. Ca^{2+} videos were acquired at 15 frames per second with 66.56-ms exposure. An optimal light-emitting diode (LED) power was selected for each mouse based on the intensity of GCamp6f fluorescence in the FOV (pixel values) to prevent oversaturation of $\Delta F/F$ changes in GCamp6f fluorescence. The same LED settings were maintained for each mouse throughout the series of imaging sessions. Two subsequent cohorts were run for Ca^{2+} imaging experiments, consisting of 6–8 mice per cohort.

Ca^{2+} image processing. Image processing was performed using Mosaic software (version 1.2, Inscopix). Videos were spatially downsampled by a binning factor of 4 (16×) and lateral brain movement was corrected using the registration engine Turboreg[34,35], which uses a single reference frame and high-contrast features in the image to shift frames with motion to matching XY positions throughout the video. Black borders from XY translation after motion correction were cropped and fluorescence changes were detected by generating a $\Delta F/F_0$ video using a minimum z-projection image of the entire movie as the reference F_0 to normalize fluorescence signals to the minimum fluorescence of pixels within the frame. Videos were then temporally downsampled by a binning factor of 3 (to 5 frames per second). Putative single cells and their respective Ca^{2+} transients were isolated with an automated cell-segmentation algorithm that uses independent and principal component analyses on $\Delta F/F_0$ videos[36]. Identified putative cells were then sorted upon visual inspection to select for units with appropriate spatial configuration and GCamp6f Ca^{2+} dynamics consistent with signals from individual neurons. Ca^{2+} transient events were then defined by a Ca^{2+} event-detection algorithm that identifies large amplitude peaks with fast rise times and exponential decays (parameters: $\tau = 200$ ms, Ca^{2+} transient event minimum size = 6 median average deviation).

Ca^{2+} data analysis. Ca^{2+} transient event rates and mouse behaviour were analysed with custom functions in MATLAB (Mathworks). To calculate the rate of Ca^{2+} transients per neuron while mice were being attacked by CD1 aggressors, attack versus no-attack periods were manually scored using Observer XT software as described above. Ca^{2+} transient event rates were calculated as the total number of Ca^{2+} events during the attack epochs, divided by the total lengths of all attack epochs per recording session. The bin sizes therefore depend upon the total time for which the aggressive mouse was attacking the experimental mouse, which ranged from 62–150 s for attack epochs. Consequently, no-attack epochs ranged from 450–538 s (that is, 600 s − attack epochs). Occupancy in different arena zones in the social interaction test (social interaction zone versus corner zones), or in the open field test was defined using EthoVision XT software. Ca^{2+} transient rates were calculated as the total number of Ca^{2+} events during exploration of the social interaction zone, corner zone and open field, and divided by the total lengths of the respective epochs per recording session. All behavioural data were exported as a logical output at 30 frames per second. Behaviour was downsampled to 5 frames per second to match Ca^{2+} transient data sampling. Logical indexing was used to calculate Ca^{2+} transient rates in different behavioural conditions (that is, attack versus no-attack, social interaction zone versus corner zones). Ca^{2+} transient rates for distances away from the centre in the open field arena were calculated as the mean firing rate per position for each individual time frame. The area under significant Ca^{2+} transients (AUC) was obtained from a subset of cells by applying a moving average smoothing filter to calculate the baseline Ca^{2+} activity during each 10 min defeat recording session. The area between the Ca^{2+} activity trace and the smoothed baseline was then calculated from the beginning to the end of each significant Ca^{2+} transient. AUC rates were calculated by summing AUC values from all significant events within a specific behavioural epoch (that is, attack or no-attack) and by dividing these summed AUC values by the total time of the respective behavioural epoch (that is, by the total lengths of all attack or no-attack periods). Amplitude rates were calculated by dividing the total summed amplitude of all significant Ca^{2+} transient events per cell within each behavioural epoch by the total length of the behavioural epoch.

Cell selectivity analysis. For defining cell selectivity, Ca^{2+} events were shuffled in time for individual cells (1,000 iterations). Shuffled rates were recalculated with logical behavioural indexing to generate null distributions of condition or zone Ca^{2+} transient event rates for each cell. A cell was considered selective for a condition or zone if its Ca^{2+} transient event rate difference between conditions or zones exceeded a 1.5-standard-deviation threshold from the null distribution (during defeat: attack versus no-attack; social interaction test: social interaction zone versus corner zones versus not in any zone; open field: centre versus periphery). Cells with Ca^{2+} transient event rate ratios at the boundary between the tentative non-selective area and the tentative zone-selective area were fitted using a linear regression to produce a population threshold line. The smallest shift of the coefficients in the linear population threshold line that resulted in grouping non-selective cells with no aggregate zone preference was chosen. For cell-selectivity analysis in the social interaction test, if a mouse did not enter either the social interaction zone or the corner zone then it was excluded from the selectivity analysis because one could not determine whether a cell was exclusively selective for either zone.

In vitro electrophysiology. One hour after the last social defeat session, mice were anaesthetized by isoflurane inhalation, decapitated, and brains rapidly removed. DG coronal slices (350 μm) were cut on a vibratome (Leica VT1000S) in ice-cold partial sucrose artificial cerebrospinal fluid (aCSF) solution (in mM): 80 NaCl, 3.5 KCl, 4.5 $MgSO_4$, 0.5 $CaCl_2$, 1.25 H_2PO_4, 25 $NaHCO_3$, 10 glucose, and 90 sucrose equilibrated with 95% O_2/5% CO_2 and stored in the same solution at 37 °C for 30 min, then at room temperature until use. Recordings were made at 30–32 °C (TC324-B; Warner Instrument Corp.) in aCSF (in mM: 124 NaCl, 2.5 KCl, 1 NaH_2PO_4, 25 $NaHCO_3$, 20 glucose, 1 $MgCl_2$, 2 $CaCl_2$). Whole-cell recordings (−70 mV) were obtained using a patch pipette (4.5–6.5 mOhm) containing (in mM): 135 KMethylsulfate, 5 KCl, 0.1 EGTA-Na, 10 HEPES, 2 NaCl, 5 ATP, 0.4 GTP, 10 phosphocreatine, 0.2% biocytin (pH 7.2; 280–290 mOsm). Patch pipettes were made from borosilicate glass (A-M Systems) using a micropipette puller (Model P-1000; Sutter Instruments). Recordings were made without correction for junction potentials. Granule cells were visualized and targeted via infrared-differential interference contrast (IR-DIC; 40× objective) optics on an Axioskop-2 FS (Zeiss). Adult-born and mature granule cells were differentiated on the basis of somatic location, input resistance and spike patterns[37,38]. Post hoc analysis of biocytin fills was used to confirm cell identity (see below). For experiments requiring perforant path stimulation, a concentric bipolar stimulating electrode (FHC) controlled by a stimulus isolator (ISO-flex, AMPI Instruments) was positioned on the DG molecular layer (triggered at 0.03 Hz). Perforant path inputs were stimulated at 40 Hz (20 pulses every 25 ms). Current and voltage signals were recorded with a MultiClamp 700B amplifier (Molecular Devices), digitized at 5–10 kHz, and filtered at 2.5–4 kHz. Data were acquired and analysed using Axograph (Axograph Scientific). Evoked synaptic responses were quantified by calculating the AUC (cumulative area above (+) and below (−) the baseline in mV s).

Immunohistochemistry. *Perfusions and sectioning.* One hour after the last social defeat session, mice were anaesthetized with ketamine/xylazine (100 mg ml^{-1} ketamine, 20 mg ml^{-1} xylazine) and transcardially perfused with 30 ml cold saline, followed by 30 ml ice-cold 4% paraformaldehyde in saline. Brains were dissected and post-fixed overnight in 4% paraformaldehyde at 4 °C. Brains were then cryoprotected in 30% sucrose (wt/vol) with 0.02% NaN_3 (wt/vol) for two days at 4 °C and subsequently frozen in optimum cutting temperature (OCT) compound (Tissue Tek) and stored at −80 °C until cryostat sectioning. Serial sections (35 μM) were cut through the entire hippocampus on a cryostat (Leica CM3050S) and stored in phosphate-buffered saline (PBS) with 0.02% NaN_3 until further processing by immunohistochemistry.

Immunohistochemistry. Sections were washed three times (10 min each time) in PBS and 0.3% Triton X-100 (PBST), treated with 1% H_2O_2 (wt/vol) in 1:1 PBS and methanol for 15 min to quench endogenous peroxidase activity, incubated in 10% normal donkey serum (NDS; wt/vol) in PBST for 2 h, and then incubated overnight at 4 °C in primary antibody in PBST + 10% NDS (goat anti-Dcx, C-18, 1:500; rabbit anti-c-fos, sc-52, 1:500, both from Santa Cruz Biotechnology; rabbit anti-mCherry,

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

1:500, ab167453; chicken anti-eYFP (eYFP, enhanced yellow fluorescent protein), 1:500, ab13970, both from Abcam; Cy2 streptavidin, 1:500, 016-160-084, Jackson Immunoresearch, to label biocytin-filled cells). The next day, sections were washed three times (10 min each time) in PBST and incubated in secondary antibody (Cy3 donkey anti-goat, 1:500, 705-165-147; Cy2 donkey anti-rabbit, 1:500, 711-225-152; Cy3 donkey anti-rabbit, 1:500, 711-165-152; biotinylated donkey anti-chicken, 1:250, 703-065-155; all from Jackson ImmunoResearch) in PBST + 10% NDS for 2 h at 21–26 °C. For eYFP stainings, sections were washed three times (10 min each time) in PBST and incubated in tertiary antibody (Cy2 streptavidin, 1:500, 016-220-084, Jackson) in PBST + 10% NDS for 2 h at 21–26 °C. All sections were then washed three times (10 min each time) in PBST, incubated in Hoechst33342 dye (1:10,000) in PBS for 5 min, and washed again three times (10 min each time) in PBS. Sections were then mounted onto glass slides and coverslipped with Aqua Poly/Mount (PolySciences) for microscopic analysis.

BrdU administration and immunohistochemistry. *iBax* mice were administered BrdU (150 mg kg^{-1}, i.p., dissolved in saline) on two consecutive days, beginning 48 h after the last tamoxifen injection. BrdU-injected mice underwent chronic social defeat stress (10 days) and were perfused following the last social defeat session on day 10. For BrdU immunohistochemistry, sections were washed three times (10 min each time) in TBS and then exposed to citrate buffer (10 mM citric acid, pH 6.0 at 95 °C) for 2 h. After three 10-min washes in TBS, sections were incubated for 2 h in 10% normal donkey serum and then incubated overnight at 4 °C with primary antibody in TBS (anti-BrdU rat monoclonal, BU1/75 (ICR1), Serotec, 1:100). The next day, sections were washed three times (10 min each time) with TBS and incubated for 2 h with secondary antibody (Cy3-conjugated goat anti-rat, A10522, Molecular Probes, 1:500) before processing for microscopy as described above.

Cell counting. Immunoreactive cells (c-fos, Dcx, BrdU) were counted by an experimenter that was blind to the experimental conditions on a Zeiss Axioplan-2 upright microscope. For each mouse, both hemispheres were counted on every sixth section along the longitudinal axis of the hippocampus (12 sections in total). Cell counts were averaged across both hemispheres for each section. To assess lateral septum activation, c-fos was counted in both hemispheres at bregma level 0.14 mm.

Confocal microscopy. To identify co-localization of mCherry and biocytin, z-stacks of immunolabelled sections were obtained using a confocal scanning microscope (Leica TCS SP8, Leica Microsystems Inc.) equipped with three simultaneous PMT detectors. Fluorescence from Cy2 was excited at 488 nm and detected at 505–550 nm, and fluorescence from Cy3 was excited at 552 nm and detected at 600–650 nm. For 488- and 552-nm excitation, the beam path included a TD 488/552/638 beam-splitter. All z-stacks were imaged with a dry Leica 20× objective (numerical aperture 0.70, working distance 0.5 mm), with a FOV of 553.6 μm × 553.6 μm, a pixel size of 0.54 × 0.54 μm, optical sectioning of 2.36 μm, a z-step of 3 μm, and a total z-thickness of 60–80 μm. Acquisition parameters included a speed of 400 Hz, pixel dwell time of 600 ns, and no line or frame averaging, resulting in a frame rate of 0.15 frames per second. Z-stacks of the area of interest were obtained using LAS X software (Leica Application Suite X 3.1.1.15751).

Statistics. Behavioural data, Dcx expression, BrdU labelling and excitatory postsynaptic potential charges were analysed using two-way analysis of variance (ANOVA) to assess genotype (Cre^- versus Cre^+) and stress effects (control versus defeated), or to assess treatment (vehicle versus CNO) and stress effects (control versus defeated), or to assess irradiation (sham versus X-ray) and genotype (Cre^- versus Cre^+) effects. Fisher's least significant difference post hoc test was used to compare individual groups when significant interactions were found. Planned comparison *t*-tests were used where appropriate when main effects were significant without significant interactions. Repeated-measures two-way ANOVA was used to assess c-fos expression. Paired two-tailed Student's *t*-tests were used to compare changes in excitatory postsynaptic potential charge in Cre^- and Cre^+ *i-hM$_4$D$_i$* mice, and changes in membrane potential in AAV8-CamKII-hM$_4$D$_i$ and AAV8-CamkII-hM3Dq infected cells (all before and after bath application of CNO). Paired one-tailed Student's *t*-tests were used to compare changes in spiking, input resistance and membrane potential of abGCs upon bath application of CNO. The unpaired two-tailed Student's *t*-test was used to assess differences in attack duration and number of attacks per mouse. The unpaired two-tailed Student's *t*-test was also used to compare behavioural data of Cre^- mice and Cre^+ *iBax* mice after social defeat in in vivo imaging experiments. Ca^{2+} transient rate data was analysed using the Mann–Whitney *t*-test (unpaired comparison) and the Wilcoxon matched pairs signed rank test (paired comparison). The chi-squared test of proportions was used to assess changes in cell selectivity from day 1 to day 10 of defeat. The Kruskal–Wallis one-way ANOVA with Dunn's post hoc test was used to assess Ca^{2+} transient rate changes for 5-cm distance bins away from the centre point of the open field. The two-sample Kolmogorov–Smirnov test was used to assess differences in cumulative frequency distributions of Ca^{2+} data between Cre^- and Cre^+ *iBax* mice.

Data was confirmed to be normally distributed using the Shapiro–Wilk test except where noted. Ca^{2+} transient rate data was not normally distributed. All statistical analyses were carried out using GraphPad Prism Software (version 6.0). Data was considered significant if $P < 0.05$, except when multiple comparisons were performed on non-parametric data, in which case α was set to $0.05/n$ where n was the number of hypotheses tested.

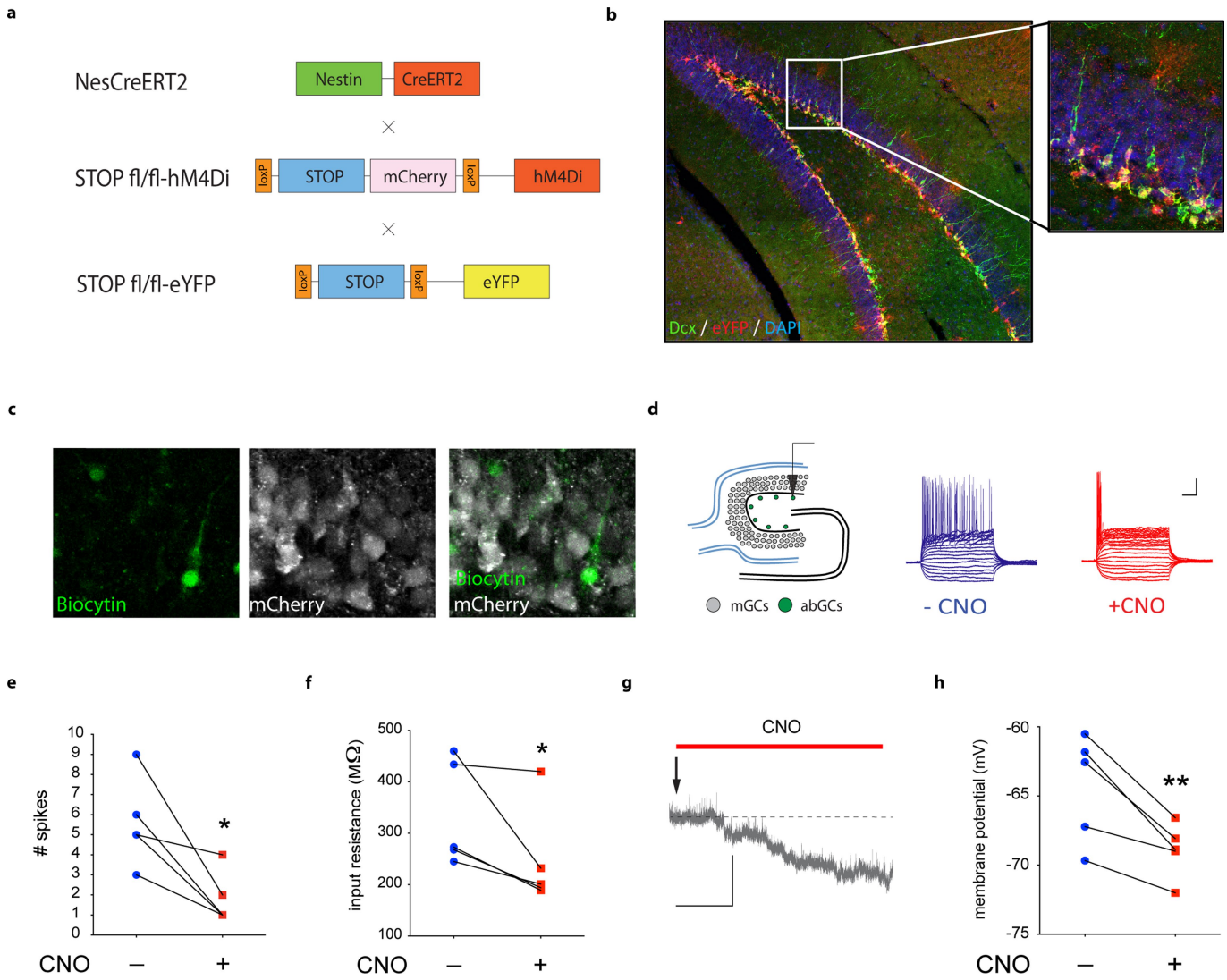
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The source data supporting the findings of this study are included in this manuscript. Additional source data are available from the corresponding authors upon reasonable request.

Code availability. Custom MATLAB scripts are available from the corresponding authors upon reasonable request.

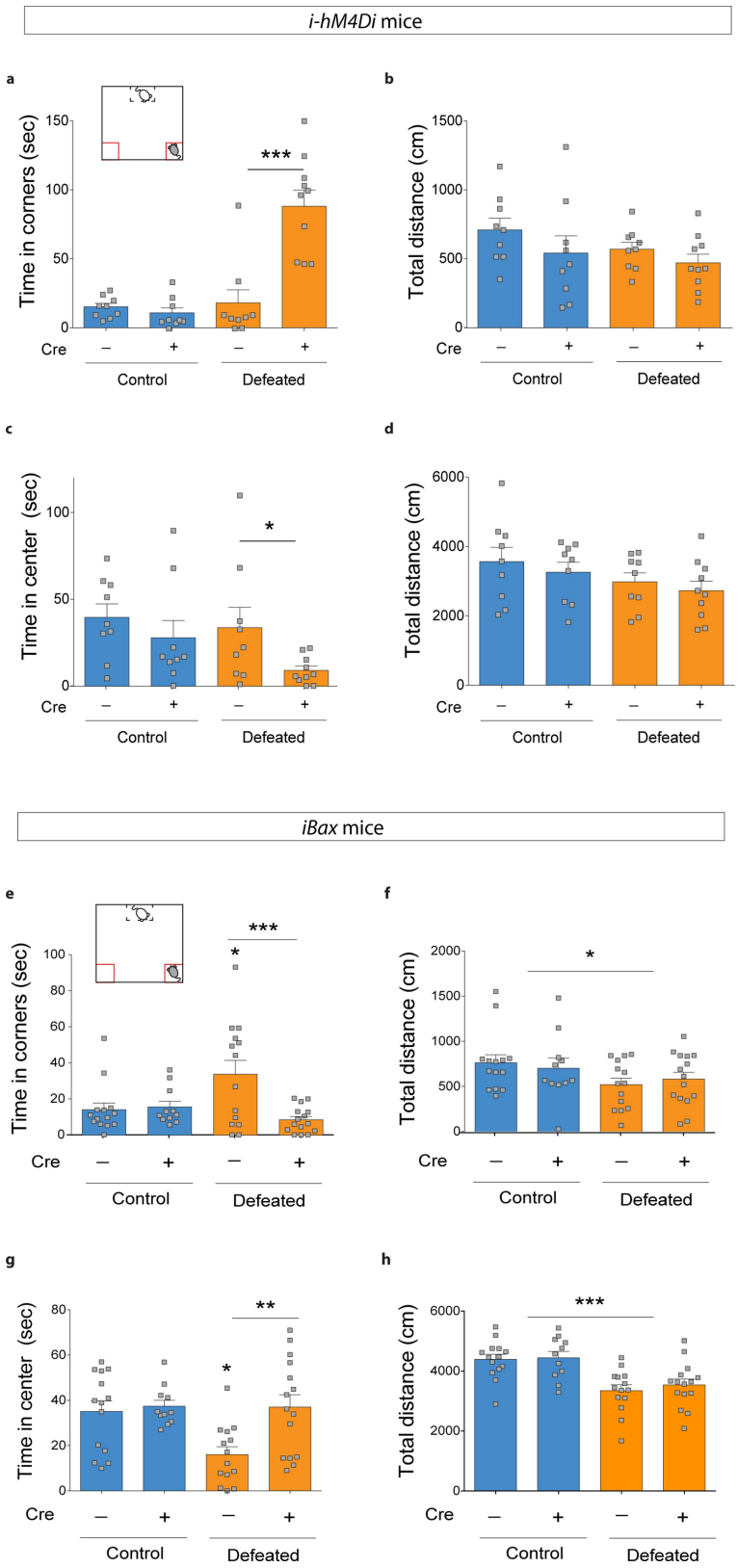
30. Golden, S. A., Covington, H. E., III, Berton, O. & Russo, S. J. A standardized protocol for repeated social defeat stress in mice. *Nat. Protoc.* **6**, 1183–1191 (2011).
31. Anacker, C. et al. Neuroanatomic differences associated with stress susceptibility and resilience. *Biol. Psychiatry* **79**, 840–849 (2016).
32. Resendez, S. L. et al. Visualization of cortical, subcortical and deep brain neural circuit dynamics during naturalistic mammalian behavior with head-mounted microscopes and chronically implanted lenses. *Nat. Protoc.* **11**, 566–597 (2016).
33. Wu, M. V. & Hen, R. Functional dissociation of adult-born neurons along the dorsoventral axis of the dentate gyrus. *Hippocampus* **24**, 751–761 (2014).
34. Ghosh, K. K. et al. Miniaturized integration of a fluorescence microscope. *Nat. Methods* **8**, 871–878 (2011).
35. Ziv, Y. et al. Long-term dynamics of CA1 hippocampal place codes. *Nat. Neurosci.* **16**, 264–266 (2013).
36. Mukamel, E. A., Nimmerjahn, A. & Schnitzer, M. J. Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron* **63**, 747–760 (2009).
37. Mongiat, L. A., Espósito, M. S., Lombardi, G. & Schinder, A. F. Reliable activation of immature neurons in the adult hippocampus. *PLoS ONE* **4**, e5320 (2009).
38. Dieni, C. V., Nietz, A. K., Panichi, R., Wadiche, J. I. & Overstreet-Wadiche, L. Distinct determinants of sparse activation during granule cell maturation. *J. Neurosci.* **33**, 19131–19142 (2013).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 1 | Silencing adult-born neurons in *i-hM$_4$D$_i$* mice. **a**, Mice expressing a TMX-inducible CreER^{T2} recombinase expressed under the control of the Nestin gene promoter[17] were crossed to mice expressing *loxP*-STOP-mCherry-*loxP*-hM$_4$D$_i$ under the control of the CAG promoter[18]. The eYFP reporter gene was used to visualize recombination by crossing *Nestin-CreERT2*$^{+/-}$;*loxP-stop-mCherry-loxP-hM$_4$D$_i$* mice to *loxP-stop-loxP-eYFP* mice (*Nestin-CreERT2*$^{+/-}$;*stop-mCherry*$^{fl/-}$*-hM$_4$D$_i$;stop*$^{fl/-}$*-eYFP*). **b**, Cre-mediated recombination in Dcx-positive adult-born neurons was visualized using eYFP/ Dcx co-labelling. DAPI, 4′,6-diamidino-2-phenylindole; eYFP, enhanced yellow fluorescent protein. **c**, Biocytin was injected into abGCs during in vitro whole-cell recordings. Expression of hM$_4$D$_i$ in recorded abGCs was confirmed by biocytin$^+$/mCherry$^-$ immunofluorescence. **d**, Schematic illustrating 500-ms current-step injections into abGCs. Representative recordings from abGCs before and after bath application of CNO (5 μM) are shown. Scale bars: 20 mV, 100 ms. **e**, Bath application of CNO decreases spiking of hM$_4$D$_i$-expressing abGCs (paired, two-tailed *t*-test, $^*P = 0.024$, $n = 5$). **f**, CNO decreases input resistance of abGCs (paired, one-tailed *t*-test, $^*P = 0.037$, $n = 5$). **g**, Representative resting membrane potential trace of abGCs upon bath application of CNO. Scale bars: 2 mV, 100 ms. **h**, CNO decreases resting membrane potential of abGCs (paired, one-tailed *t*-test, $^{**}P = 0.006$, $n = 5$); Scale bars: 5 mV, 1 min. Error bars, ±s.e.m.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

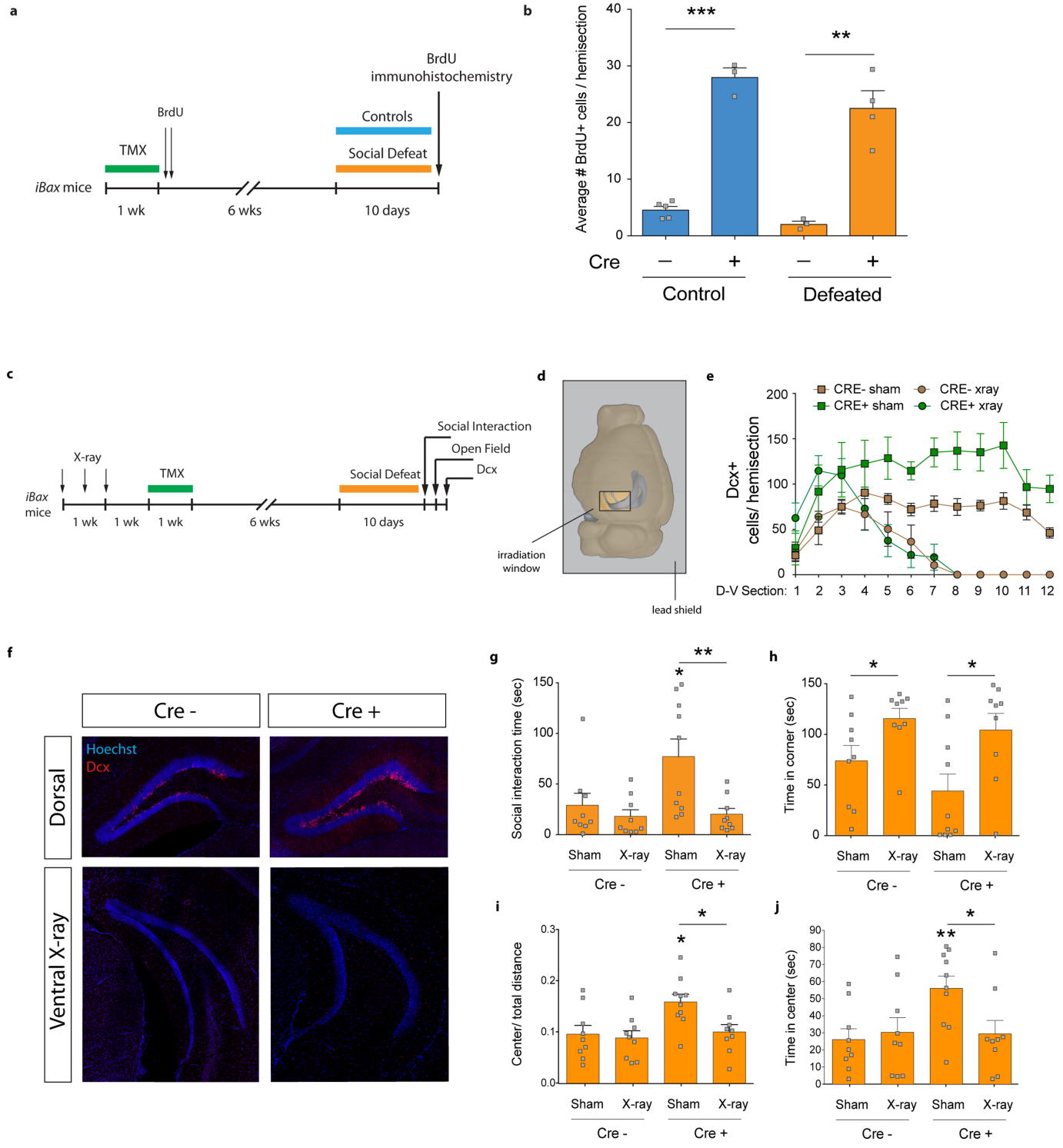


Extended Data Fig. 2 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 2 | *i-hM$_4$D$_i$* mice and *iBax* mice behaviour: time in corner of the social interaction test and time in centre of the open field. **a**–**d**, *i-hM$_4$D$_i$* mice. **a**, Silencing abGCs in Cre$^+$ *i-hM$_4$D$_i$* mice increased the time spent in the corner of the social interaction test after subthreshold defeat (interaction $F_{1,33} = 23.0$, ***$P < 0.0001$, genotype effect $F_{1,33} = 17.96$, ***$P = 0.0002$, stress effect $F_{1,33} = 26.7$, ***$P < 0.0001$, post hoc test, ***$P < 0.0001$). **b**, No difference in total distance travelled in the social interaction test (interaction $F_{1,33} = 0.17$, $P = 0.68$; genotype effect $F_{1,33} = 2.5$, $P = 0.12$; stress effect $F_{1,33} = 1.6$, $P = 0.22$) **c**, Time spent in the centre of the open field after subthreshold defeat (interaction $F_{1,33} = 0.57$, $P = 0.46$, genotype effect $F_{1,33} = 4.63$, *$P = 0.04$, stress effect $F_{1,33} = 2.14$, $P = 0.15$; planned comparison *t*-test: Cre$^-$ defeat versus Cre$^+$ defeat, *$P = 0.045$). **d**, No difference in total distance travelled in the open field (interaction $F_{1,33} = 0.007$, $P = 0.93$; genotype effect $F_{1,33} = 0.79$, $P = 0.38$; stress effect $F_{1,33} = 3.22$, $P = 0.08$). $n = 9, 9, 9, 10$. **e**–**h**, *iBax* mice. **e**, Chronic defeat (10 days) increased the amount of time that Cre$^-$ mice spent in the corner during the social interaction test. This effect was absent in Cre$^+$ *iBax* mice (interaction $F_{1,50} = 7.9$, **$P = 0.007$; genotype effect $F_{1,50} = 6.3$, *$P = 0.016$; stress effect $F_{1,50} = 1.7$, $P = 0.19$; post hoc test, ***$P = 0.0003$). **f**, Chronic defeat decreased the total distance travelled in the social interaction test without genotype effects (interaction $F_{1,50} = 0.5$, $P = 0.48$; genotype effect $F_{1,50} = 0.00001$, $P = 0.99$; stress effect $F_{1,50} = 4.4$, *$P = 0.04$). **g**, Chronic defeat decreased the time Cre$^-$ mice spent in the centre of the open field. This effect was absent in Cre$^+$ *iBax* mice (interaction $F_{1,50} = 4.5$, *$P = 0.039$; genotype effect, $F_{1,50} = 6.9$, *$P = 0.01$; stress effect, $F_{1,50} = 4.8$, *$P = 0.03$; post hoc test, **$P = 0.001$). **h**, Chronic defeat decreased the total distance travelled in the open field without genotype effects (interaction $F_{1,50} = 0.12$, $P = 0.73$; genotype effect, $F_{1,50} = 0.39$, $P = 0.54$; stress effect $F_{1,50} = 25.2$, ***$P < 0.0001$). $n = 14, 11, 14, 15$. Error bars, ±s.e.m.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

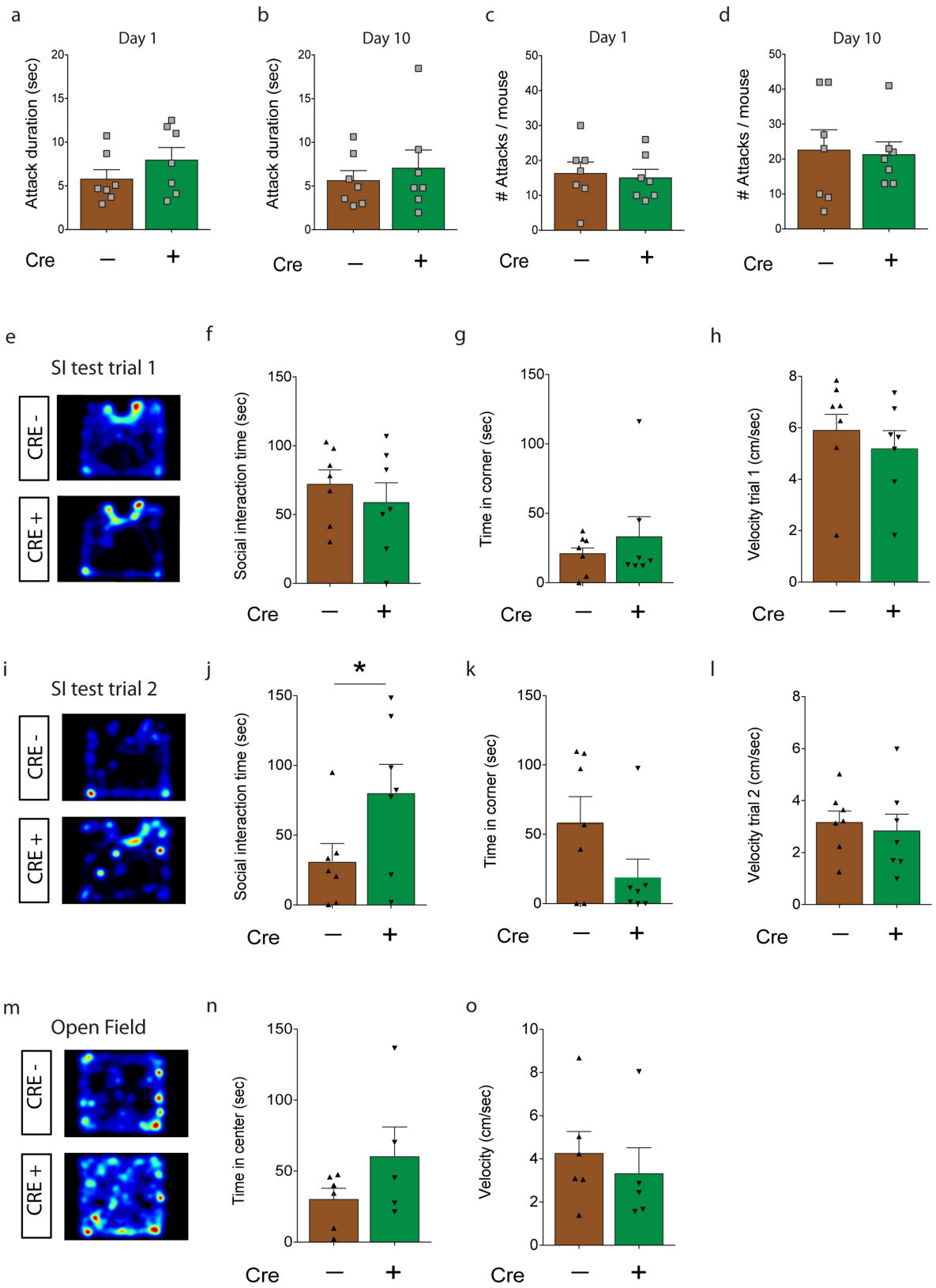


Extended Data Fig. 3 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 3 | *iBax* mice show higher levels of adult-born cell survival and ventral X-ray irradiation abolishes stress resilience. **a**, Experimental timeline for cell survival assessment using BrdU incorporation. **b**, The number of cells that have incorporated BrdU at 8 weeks after BrdU injection is increased in undefeated mice and in defeated Cre^+ *iBax* mice compared to Cre^- control mice (interaction $F_{1,11} = 0.69$, $P = 0.43$; genotype effect $F_{1,11} = 138.9$, $^{***}P < 0.0001$; stress effect $F_{1,11} = 4.8$, $P = 0.05$; planned comparison *t*-test, control Cre^- versus Cre^+, $^{***}P < 0{\cdot}001$; defeat Cre^- versus Cre^+, $^{**}P = 0{\cdot}002$; $n = 5, 3, 3, 4$). **c**, Experimental timeline for X-ray irradiation. **d**, Schematic indicating sliding window position of a protective lead shield during X-ray irradiation of the vDG. **e**, Ventral X-ray irradiation eliminated Dcx^+ cells in the vDG in Cre^- and Cre^+ mice. $n_{Cre^- sham} = 5$, $n_{Cre+sham} = 4$, $n_{Cre-Xray} = 5$, $n_{Cre+Xray} = 5$. **f**, Representative images of Dcx expression in the dorsal dentate gyrus and the ventral dentate gyrus in X-ray irradiated mice. **g**, Cre^+ *iBax* mice showed longer social interaction times than Cre^- mice after chronic defeat. This effect was abolished by ventral X-ray irradiation (interaction $F_{1,33} = 3.8$, $P = 0.06$; genotype effect $F_{1,33} = 4.5$, $^{*}P = 0.04$; X-ray effect, $F_{1,33} = 8.4$, $^{**}P = 0.007$; planned comparison *t*-test, Cre^- sham versus Cre^+ sham, $^{*}P = 0.04$; Cre^+ sham versus Cre^+ X-ray, $^{**}P = 0.008$; $n = 9, 9, 10, 9$). **h**, Ve*n*tral X-ray irradiatio*n* increased time spent in the corner in both genotypes (interaction $F_{1,33} = 0.38$, $P = 0.54$; genotype effect, $F_{1,33} = 1.8$, $P = 0.18$; X-ray effect, $F_{1,33} = 11.6$, $^{**}P = 0.0018$; planned comparison *t*-test, Cre^- sham versus Cre^- X-ray, $^{*}P = 0.04$; Cre^+ sham versus Cre^+ X-ray, $^{*}P = 0.02$; $n = 9, 9, 10, 9$). **i**, Cre^+ *iBax* mice show increased centre/total distance in the open field compared with Cre^- mice. Ventral X-ray irradiation abolished this effect (interaction $F_{1,33} = 2.9$, $P = 0.1$; genotype effect, $F_{1,33} = 6.1$, $^{*}P = 0.019$; X-ray effect, $F_{1,33} = 4.8$, $^{*}P = 0.036$; planned comparison *t*-test, Cre^- sham versus Cre^+ sham, $^{*}P = 0.01$; Cre^+ sham versus Cre^+ X-ray, $^{*}P = 0.01$; $n = 9, 9, 10, 9$). **j**, Cre^+ *iBax* mice show increased time spent in the centre of the open field than Cre^- mice. Ventral X-ray irradiation abolished this effect (interaction $F_{1,33} = 4.2$, $^{*}P = 0.049$; genotype effect $F_{1,33} = 3.7$, $P = 0.06$; X-ray effect, $F_{1,33} = 2.2$, $P = 0.15$; post hoc test, Cre − sham versus Cre + sham, $^{**}P = 0.008$; Cre + sham versus Cre + X-ray, $^{*}P = 0.02$; $n = 9, 9, 10, 9$). Error bars, ±s.e.m.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

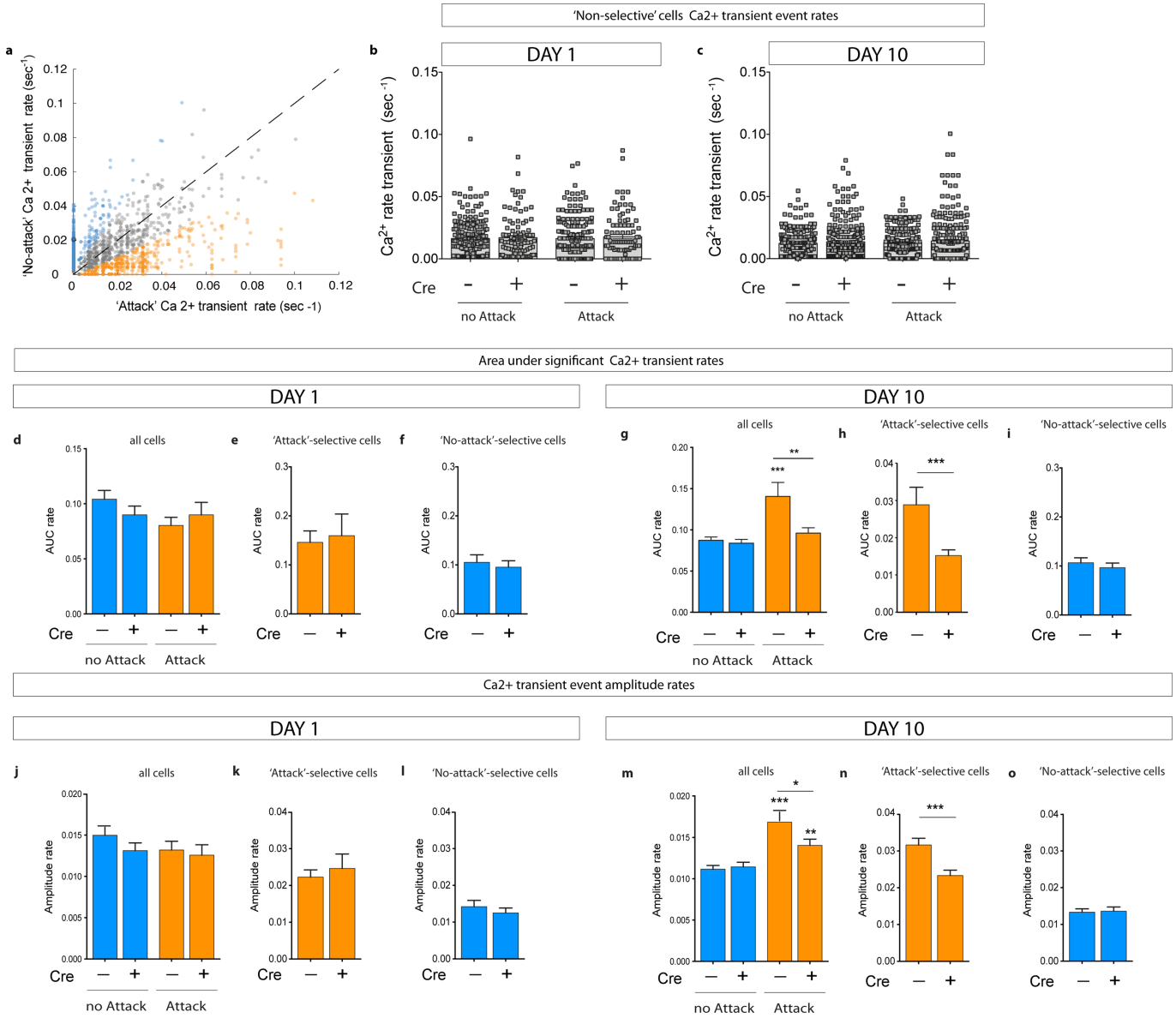


Extended Data Fig. 4 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

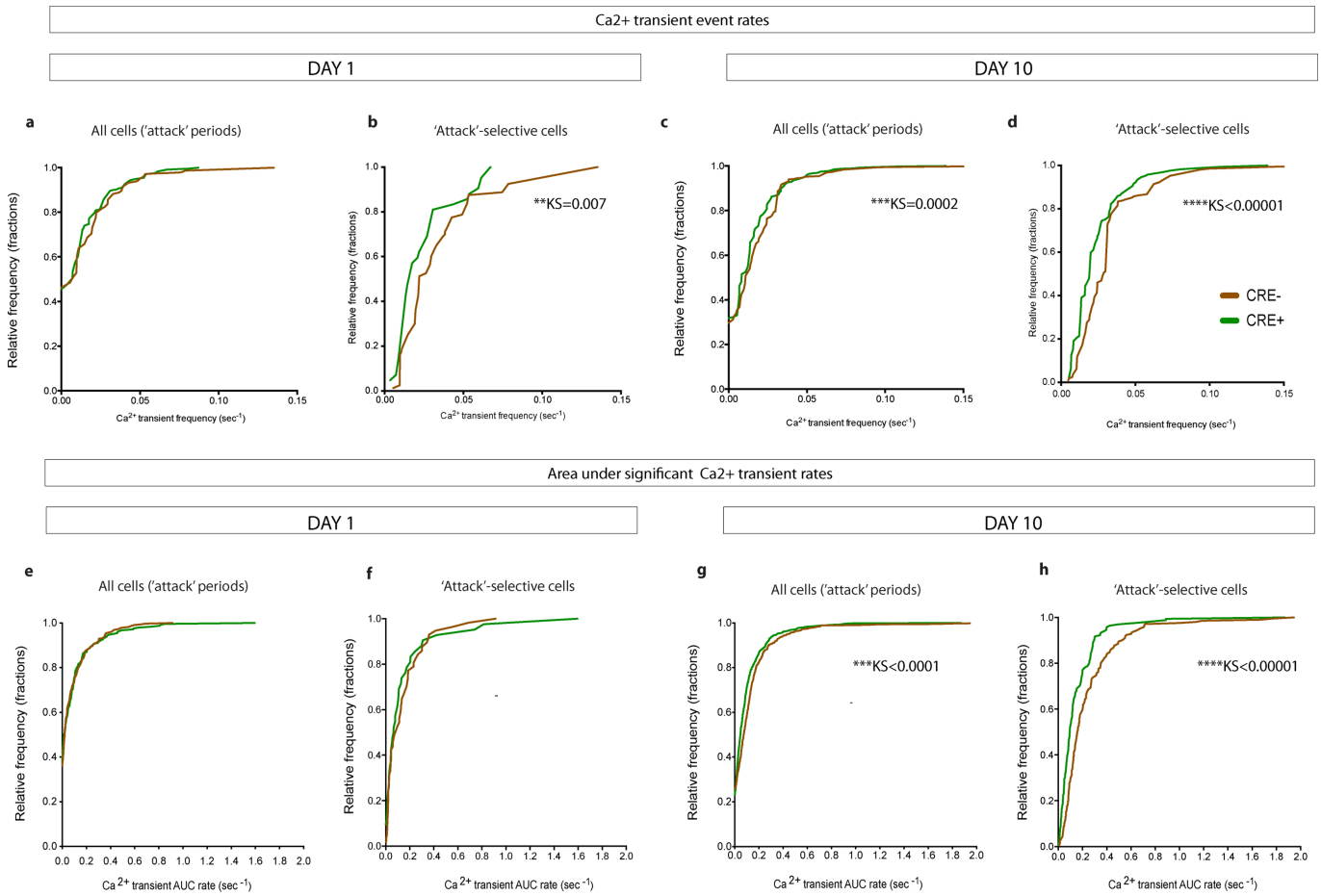
Extended Data Fig. 4 | *iBax* mice: Number of attacks and length of attacks during social defeat, and behaviour after 10 days of chronic social defeat for all imaged mice. a, b, No difference in average duration of attacks on the first day of social defeat (**a**; unpaired, two-tailed *t*-test, $P = 0.25$; $n = 7, 7$), or the last day of social defeat (**b**; $P = 0.56$; $n = 7, 7$). **c**, **d**, No difference in the average number of attacks on the first day of social defeat (**c**; $P = 0.76$; $n = 7, 7$), or the last day of social defeat (**d**; $P = 0.85$; $n = 7, 7$). **e**, Representative heat map of time spent in the social interaction test during the first trial (without CD1 mouse). **f**, No differences in time in the interaction zone during the first trial (unpaired, two-tailed *t*-test, $P = 0.47$; $n = 7, 7$). **g**, No differences in time in the corner zone during the first trial ($P = 0.46$; $n = 7, 7$). **h**, No differences in velocity during the first trial ($P = 0.43$; $n = 7, 7$). **i**, Representative heat map of time spent in the social interaction test during the second trial (with CD1). **j**, Cre^+ *iBax* mice spent longer in the interaction zone during the second trial than Cre^- mice ($^*P = 0.049$; $n = 7, 7$). **k**, Cre^+ *iBax* mice spent less time in the corner zone during the second trial than Cre^- mice ($P = 0.1$; $n = 7, 7$). **l**, No differences in velocity during the second trial ($P = 0.65$; $n = 7, 7$). **m**, Representative heat map of time spent in the open field test. **n**, Time spent in the centre of the open field ($P = 0.18$; $n_{Cre^-} = 6, n_{Cre^+} = 5$). **o**, No differences in velocity in the open field ($P = 0.56$; $n_{Cre^-} = 6, n_{Cre^+} = 5$). Error bars, ±s.e.m.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



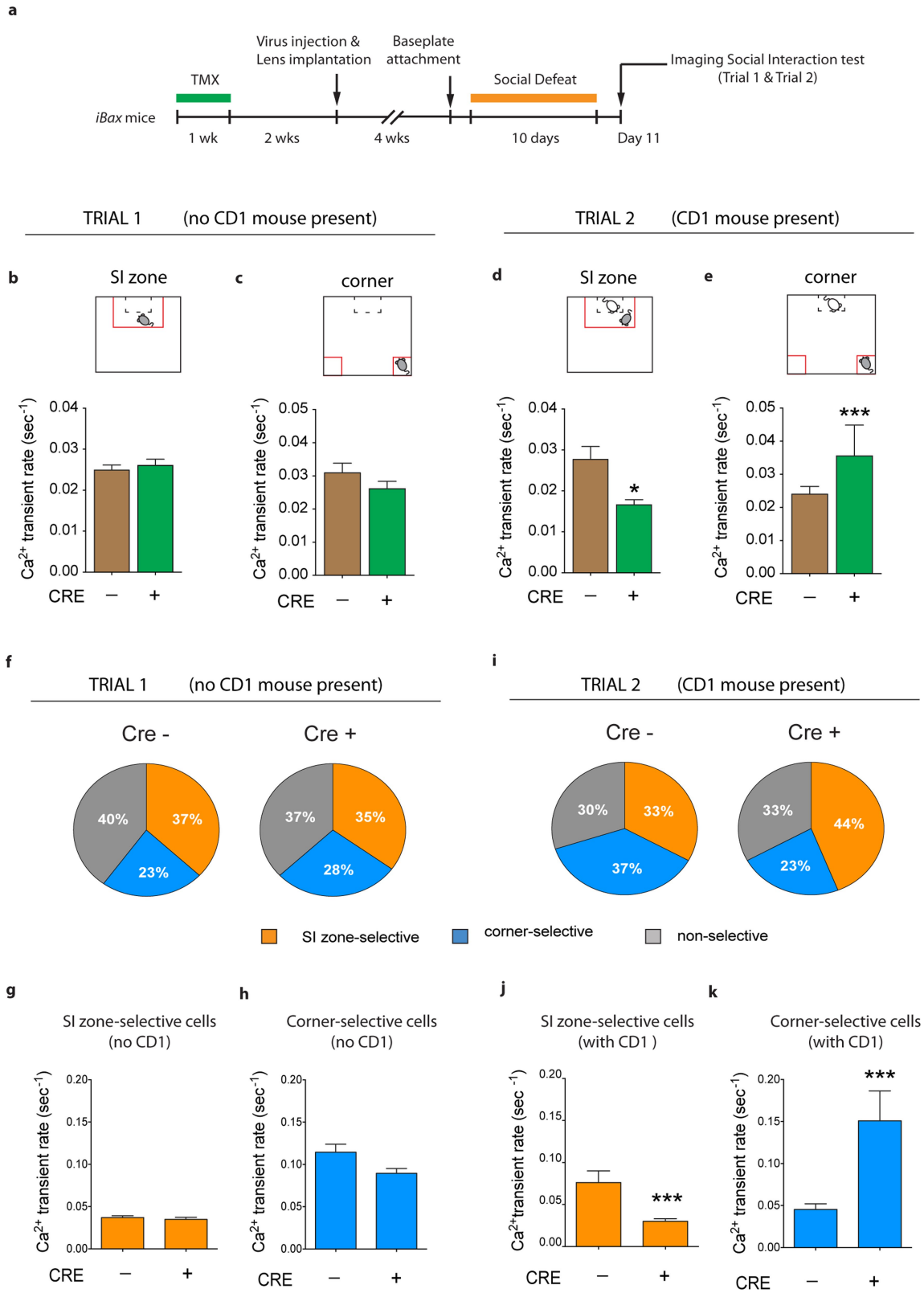
Extended Data Fig. 5 | Ca^{2+} transient rates of non-selective cells, AUC rates, and amplitude rates of significant Ca^{2+} transients on social defeat day 1 and day 10. a, Scatter plot showing the Ca^{2+} transient rates of individual neurons during attack versus no-attack periods. Colours based on selectivity for attacks (orange), no-attack (blue), and non-selectivity (grey) determined by exceeding shuffle distribution. For presentation purposes, the scatter plot shows rates up to 0.12 events per second. Ten cells showed rates greater than 0.12 events per second and were included in the cell selectivity analysis (see Source Data). **b**, **c**, No differences in Ca^{2+} transient rates of non-selective cells on day 1 (**b**) or day 10 (**c**). **d–i**, AUC rate analysis of significant Ca^{2+} transients. **d**, No differences in AUC rates between Cre^- mice and Cre^+ *iBax* mice on day 1 (no-attack: Mann–Whitney $U = 38{,}448$, $P = 0.37$; attack: $U = 40{,}070$, $P = 0.94$; $n_{Cre^-} = 327$, $n_{Cre^+} = 246$). **e**, No difference in AUC rates of attack-selective cells on day 1 ($U = 1{,}093$, $P = 0.46$; $n_{Cre^-} = 57$, $n_{Cre^+} = 42$). f, No difference in AUC rates of no-attack-selective cells on day 1 ($U = 6{,}065$, $P = 0.76$; $n_{Cre^-} = 115$, $n_{Cre^+} = 108$). **g**, AUC rates are increased only in Cre^- mice during attack periods on day 10 (Wilcoxon matched-pairs signed rank, $^{***}P_{Cre^-} = 0.0006$; $n_{Cre^-} = 620$; $P_{Cre^+} = 0.44$; $n_{Cre^+} = 592$). Cre^+ *iBax* mice show lower attack responses than Cre^- mice ($U = 165{,}092$, $^{**}P = 0.002$). **h**, AUC rates of attack-selective cells are lower in Cre^+ *iBax* mice than in Cre^- mice on day 10 ($U = 12{,}162$, $^{***}P < 0.0001$; $n_{Cre^-} = 211$, $n_{Cre^+} = 169$). **i**, No difference in AUC rates of no-attack-selective cells on day 10 ($U = 11{,}127$, $P = 0.37$; $n_{Cre^-} = 169$, $n_{Cre^+} = 140$). **j–o**, Amplitude rate analysis of significant Ca^{2+} transients. j, No differences in amplitude rates between Cre^- mice and Cre^+ *iBax* mice on day 1 (no-attack: Mann–Whitney $U = 38{,}693$, $P = 0.44$; attack: $U = 38{,}741$, $P = 0.43$; $n_{Cre^-} = 327$, $n_{Cre^+} = 246$). **k**, No difference in amplitude rates of attack-selective cells on day 1 ($U = 1{,}132$, $P = 0.65$; $n_{Cre^-} = 57$, $n_{Cre^+} = 42$). **l**, No difference in amplitude rates of no-attack-selective cells on day 1 ($U = 5{,}767$, $P = 0.36$; $n_{Cre^-} = 115$, $n_{Cre^+} = 108$). **m**, Amplitude rates are increased during attack periods on day 10 (Wilcoxon matched-pairs signed rank, $^{***}P_{Cre^-} < 0.0001$; $n_{Cre^-} = 620$; $^{**}P_{Cre^+} = 0.0019$; $n_{Cre^+} = 592$). Cre^+ *iBax* mice show lower attack responses than Cre^- mice ($U = 170{,}024$, $^{*}P = 0.0246$). **n**, Amplitude rates of attack-selective cells are lower in Cre^+ *iBax* mice than in Cre^- mice on day 10 ($U = 12{,}936$, $^{***}P < 0.0001$; $n_{Cre^-} = 211$, $n_{Cre^+} = 169$). **o**, No difference in amplitude rates of no-attack-selective cells on day 10 ($U = 11{,}518$, $P = 0.69$; $n_{Cre^-} = 169$, $n_{Cre^+} = 140$). Error bars, ±s.e.m.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 6 | Cumulative frequency distribution plots of Ca^{2+} transients. **a**, No differences in the distributions of Ca^{2+} transient event rates were observed between Cre^{-} mice and Cre^{+} *iBax* mice during attack periods on day 1. $n_{Cre^-} = 460$, $n_{Cre^+} = 246$. **b**, Distributions of Ca^{2+} transient event rates of attack-selective cells are different between Cre^{-} mice and Cre^{+} *iBax* mice on day 1. $n_{Cre^-} = 80$, $n_{Cre^+} = 42$. **c**, Distributions of Ca^{2+} transient event rates are significantly different between Cre^{-} mice and Cre^{+} *iBax* mice during attack periods on day 10. $n_{Cre^-} = 620$, $n_{Cre^+} = 592$. **d**, Distributions of Ca^{2+} transient event rates of attack-selective cells are significantly different between Cre^{-} mice and Cre^{+} *iBax* mice on day 10. $n_{Cre^-} = 211$, $n_{Cre^+} = 169$. **e**, No differences in the distributions of Ca^{2+} transient AUC rates were observed between Cre^{-} mice and Cre^{+} *iBax* mice during attack periods on day 1. $n_{Cre^-} = 327$, $n_{Cre^+} = 246$. **f**, No differences in the distributions Ca^{2+} transient AUC rates of attack-selective cells were observed between Cre^{-} mice and Cre^{+} *iBax* mice on day 1. $n_{Cre^-} = 57$, $n_{Cre^+} = 42$. **g**, Distributions of Ca^{2+} transient AUC rates are significantly different between Cre^{-} mice and Cre^{+} *iBax* mice during attack periods on day 10. $n_{Cre^-} = 620$, $n_{Cre^+} = 592$. **h**, Distributions of Ca^{2+} transient AUC rates of attack-selective cells are significantly different between Cre^{-} mice and Cre^{+} *iBax* mice on day 10. $n_{Cre^-} = 211$, $n_{Cre^+} = 169$. KS, Kolmogorov–Smirnov test *P* value (shown for significant differences only).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

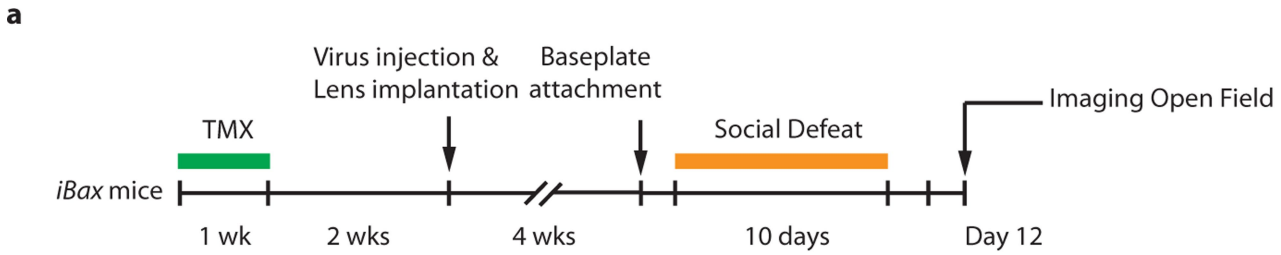


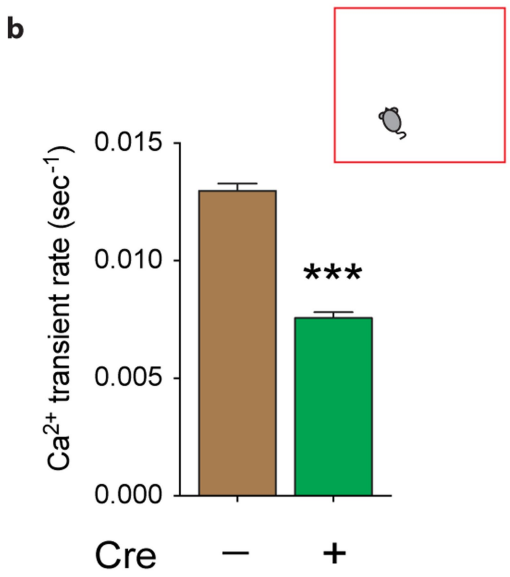
Extended Data Fig. 7 | See next page for caption.

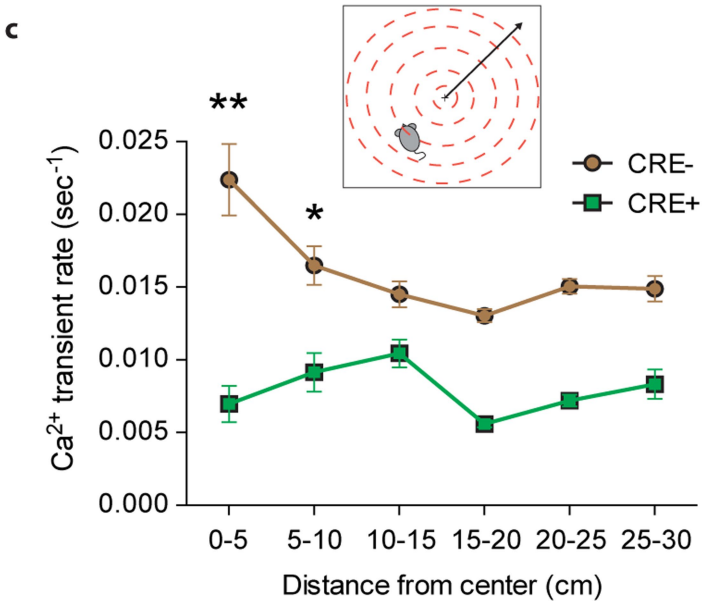
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

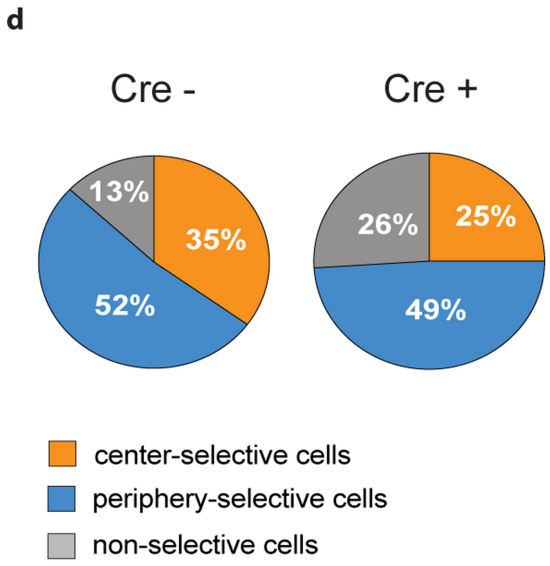
Extended Data Fig. 7 | *iBax* mice: decreased vDG activity during stress-induced social avoidance in the social interaction test. a, Experimental timeline. **b**, Ca^{2+} transient rates in the social interaction (SI) zone during trial 1 ($n_{Cre^-} = 546$, $n_{Cre^+} = 233$). **c**, Ca^{2+} transient rates in the corner zones during trial 1 ($n_{Cre^-} = 530$, $n_{Cre^+} = 329$). **d**, Cre^+ *iBax* mice show lower Ca^{2+} transient rates than Cre^- mice in the social interaction zone during trial 2 (Mann–Whitney $U = 67{,}155$, $^*P = 0.018$; $n_{Cre^-} = 483$, $n_{Cre^+} = 308$). **e**, Cre^+ *iBax* mice show higher Ca^{2+} transient rates than Cre^- mice in the corner zones during trial 2 ($U = 22{,}585$, $^{***}P < 0.0001$; $n_{Cre^-} = 301$, $n_{Cre^+} = 206$). **f**, Cell selectivity pie charts for trial 1. **g**, No differences in the Ca^{2+} transient rates of social-interaction-zone-selective cells are seen during exploration of the social interaction zone in trial 1 ($n_{Cre^-} = 187$, $n_{Cre^+} = 79$). **h**, No differences in the Ca^{2+} transient rates of corner-selective cells are seen during corner exploration in trial ($n_{Cre^-} = 119$, $n_{Cre^+} = 63$). **i**, Cell selectivity pie charts for trial 2. **j**, Social-interaction-zone-selective cells show lower Ca^{2+} transient rates in Cre^+ *iBax* mice than in Cre^- mice during exploration of the social interaction zone in trial 2 ($U = 2{,}218$, $^{***}P < 0.0001$; $n_{Cre^-} = 97$, $n_{Cre^+} = 85$). **k**, Corner-selective cells show higher Ca^{2+} transient rates in Cre^+ *iBax* mice than in Cre^- mice during corner exploration in trial 2 ($U = 1{,}465$, $^{***}P = 0.0002$; $n_{Cre^-} = 110$, $n_{Cre^+} = 43$). Error bars, ±s.e.m.

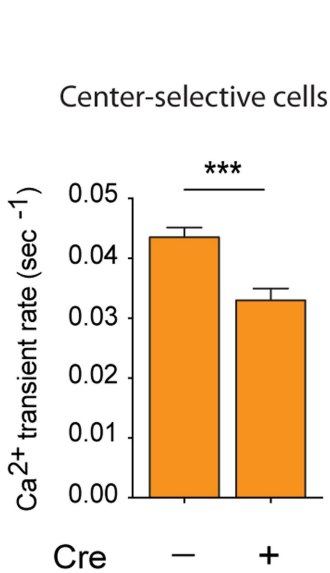
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

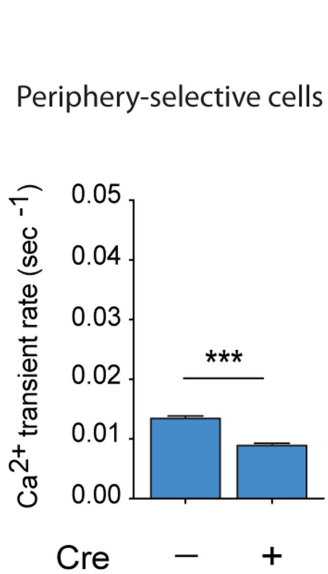






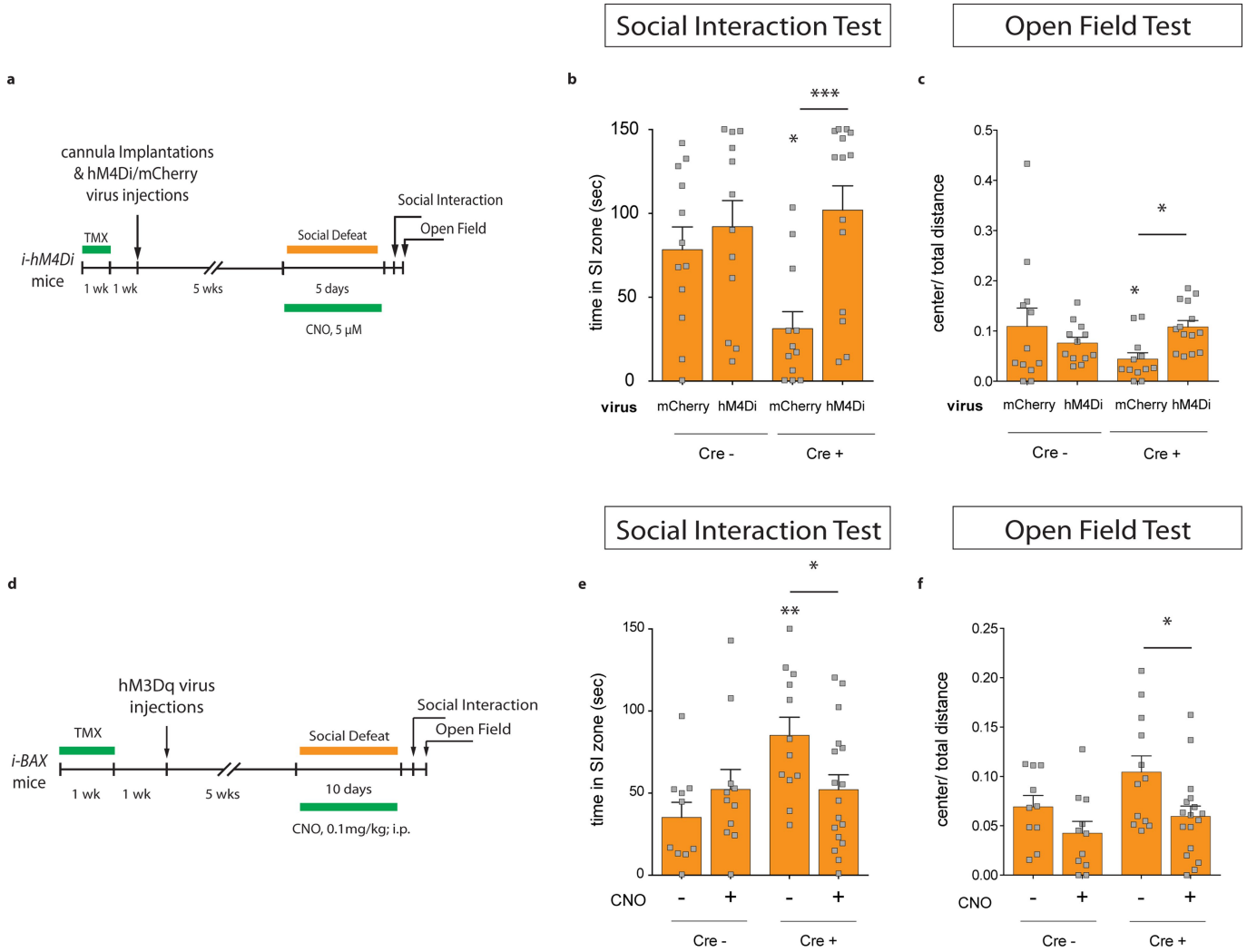






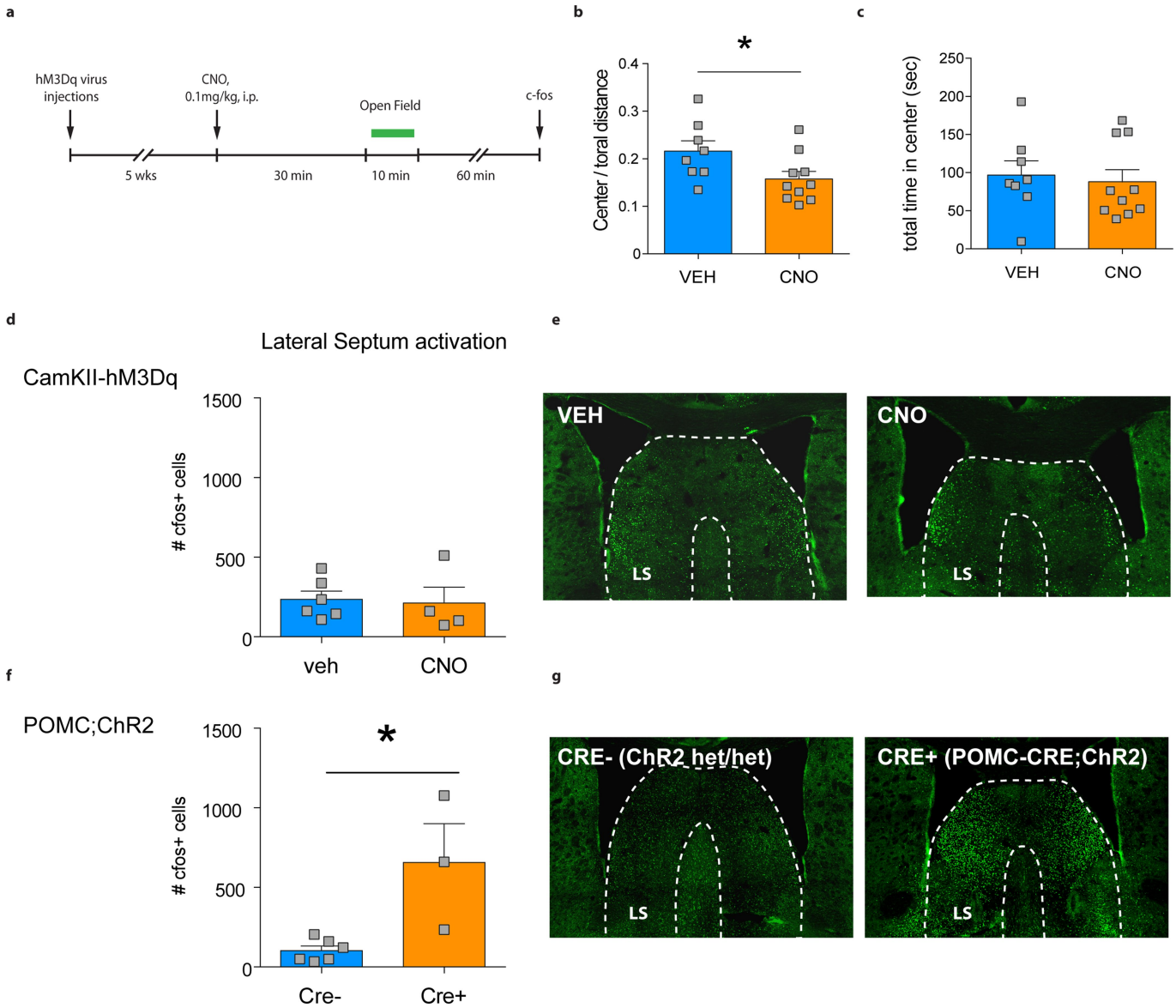
Extended Data Fig. 8 | *iBax* mice: decreased vDG activity during stress-induced anxiety-like behaviour in the open field test. a, Experimental timeline. **b**, Cre^+ *iBax* mice show lower Ca^{2+} transient rates than Cre^- mice during open field exploration (Mann–Whitney $U = 130{,}884$, $^{***}P < 0.0001$; $n_{Cre^-} = 859$ cells, $n_{Cre^+} = 494$ cells). **c**, Ca^{2+} transient rates are plotted for each 5-cm distance bin away from the centre of the open field. Cre^- mice exhibit increased Ca^{2+} transient rates in the centre compared to the periphery (Kruskal–Wallis $H = -22.57$, $^{***}P < 0.001$; Dunn's test: 0–5 cm versus 25–30 cm distance from centre point: $^{**}P = 0.002$; 5–10 cm versus 25–30 cm distance from centre point: $^{*}P = 0.026$). Cre^+ *iBax* mice did not exhibit increased Ca^{2+} transient rates in the centre compared to the periphery ($H = 0.93$, $P = 0.97$). **d**, Cell selectivity pie charts. **e**, Centre-selective cells show lower Ca^{2+} transient rates in Cre^+ *iBax* mice than in Cre^- mice during centre exploration (Mann–Whitney $U = 13{,}892$, $^{***}P < 0.001$, $n_{Cre^-} = 302$, $n_{Cre^+} = 122$). **f**, Periphery-selective cells show lower Ca^{2+} transient rates in Cre^+ *iBax* mice than in Cre^- mice during periphery exploration ($U = 34{,}033$, $^{***}P < 0.001$, $n_{Cre^-} = 450$, $n_{Cre^+} = 241$). Error bars, ±s.e.m.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 9 | Manipulating mature granule cell activity counteracts the effect of adult hippocampal neurogenesis on stress resilience. a, Experimental design for subchronic social defeat stress (5 days) while inhibiting mature granule cells in *i-hM4Di* mice. **b**, **c**, Inhibiting mature granule cells with CNO counteracts the pro-susceptibility phenotype of Cre^+ *i-hM4Di* mice in the social interaction test (**b**; interaction $F_{1,46} = 4.3$, $^*P = 0.04$; genotype $F_{1,46} = 1.8$, $P = 0.18$; CamKII-hM_4D_i $F_{1,46} = 9.5$,$^{**}P = 0.004$; post hoc test (Cre^- mCherry versus Cre^+ mCherry), $^*P = 0.02$; post hoc test (Cre^+ mCherry versus Cre^+ hM_4D_i),$^{***}P = 0.0006$; $n = 12, 12, 12, 14$) and in the open field test (**c**; interaction $F_{1,46} = 5.57$, $^*P = 0.02$; genotype $F_{1,46} = 0.64$, $P = 0.43$; CamKII-hM_4D_i $F_{1,46} = 0.56$, $P = 0.46$; post hoc test (Cre^- mCherry versus Cre^+ mCherry), $^*P = 0.033$; post hoc test (Cre^+ mCherry versus Cre^+ hM_4D_i),$^*P = 0.03$; $n = 12, 12, 12, 14$). **d**, Experimental design for chronic social defeat stress (10 days) while exciting mature granule cells in *iBax* mice. **e**, **f**, Exciting mature granule cells with CNO counteracts the pro-resilience phenotype of Cre^+ *iBax* mice in the social interaction test (**e**; interaction $F_{1,46} = 5.63$, $^*P = 0.02$; genotype $F_{1,46} = 5.52$, $^*P = 0.023$; CNO $F_{1,46} = 0.57$, $P = 0.45$; post hoc test (Cre^- vehicle versus Cre^+ vehicle), $^{**}P = 0.003$; post hoc test (Cre^+ vehicle versus Cre^+ CNO), $^*P = 0.021$; $n = 10, 11, 12, 17$) and in the open field test (**f**; interaction $F_{1,46} = 0.5$, $P = 0.49$; genotype $F_{1,46} = 4.1$, $P = 0.05$; CNO $F_{1,46} = 7.71$, $^{**}P = 0.008$; planned comparisons: unpaired *t*-test (Cre^- vehicle versus Cre^+ vehicle), $P = 0.1$; unpaired *t*-test (Cre^+ vehicle versus Cre^+ CNO), $^*P = 0.02$; $n = 10, 11, 12, 17$. Error bars, ±s.e.m.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 10 | Acute activation of vDG mature granule cells results in a modest increase in anxiety during open field exploration and does not activate the lateral septum. a, Experimental design for acute stimulation of vDG mature granule cells during open field exploration. **b**, Exciting mature granule cells with CNO during open field exploration decreases centre distance travelled as a proportion of the total distance (unpaired *t*-test; $*P = 0.039$; $n = 8, 10$). **c**, No difference was observed for the total time spent in the centre ($P = 0.72$; $n = 8, 10$). **d**, No difference in *c-fos* expression was observed in the lateral septum (LS) of mice in which the vDG was stimulated during open field exploration using CNO-mediated activation of hM_3D_q ($P = 0.82$; $n = 6, 4$). **e**, Representative images of *c-fos* expression in the lateral septum upon injection with vehicle (VEH) or CNO. **f**, Increased *c-fos* expression was observed in the lateral septum of mice upon optogenetic activation of the vDG (POMC-Cre;ChR2), which has been shown to decrease anxiety-like behaviour[8] ($*P = 0.01$; $n = 6, 3$). **g**, Representative images of *c-fos* expression in the lateral septum upon light activation of ChR2 (Cre^+; POMC-Cre;ChR2). Error bars, ±s.e.m.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

https://doi.org/10.1038/s41586-018-0226-8

A stromal cell population that inhibits adipogenesis in mammalian fat depots

Petra C. Schwalie[1,5], Hua Dong[2,5], Magda Zachara[1,5], Julie Russeil[1], Daniel Alpern[1], Nassila Akchiche[2], Christian Caprara[3], Wenfei Sun[2], Kai-Uwe Schlaudraff[4], Gianni Soldati[3], Christian Wolfrum[2]* & Bart Deplancke[1]*

Adipocyte development and differentiation have an important role in the aetiology of obesity and its co-morbidities[1,2]. Although multiple studies have investigated the adipogenic stem and precursor cells that give rise to mature adipocytes[3–14], our understanding of their in vivo origin and properties is incomplete[2,15,16]. This is partially due to the highly heterogeneous and unstructured nature of adipose tissue depots[17], which has proven difficult to molecularly dissect using classical approaches such as fluorescence-activated cell sorting and Cre–lox lines based on candidate marker genes[16,18]. Here, using the resolving power of single-cell transcriptomics[19] in a mouse model, we reveal distinct subpopulations of adipose stem and precursor cells in the stromal vascular fraction of subcutaneous adipose tissue. We identify one of these subpopulations as CD142$^+$ adipogenesis-regulatory cells, which can suppress adipocyte formation in vivo and in vitro in a paracrine manner. We show that adipogenesis-regulatory cells are refractory to adipogenesis and that they are functionally conserved in humans. Our findings point to a potentially critical role for adipogenesis-regulatory cells in modulating adipose tissue plasticity, which is linked to metabolic control, differential insulin sensitivity and type 2 diabetes.

To study the molecular characteristics and the subpopulation structure of adipogenic stem and precursor cells (ASPCs), we performed single-cell RNA sequencing (scRNA-seq) on Lin$^-$(CD31$^-$CD45$^-$TER119$^-$)CD29$^+$CD34$^+$SCA1$^+$ cells from the subcutaneous stromal vascular fraction (SVF) of transgenic mice, in which red fluorescent protein is induced in cells that express *Dlk1* (Fig. 1a and Methods). CD29, CD34 and SCA1 are expected to enrich for stem cells, and DLK1 has previously been suggested to mark pre-adipocytes[20,21]. We obtained 208 high-quality cells (Fig. 1a, Extended Data Fig. 1a–d and Methods) and confirmed that the expression of all fluorescence-activated cell sorting (FACS) markers corresponded to the enrichment strategy that was used (Extended Data Fig. 1e). The cells grouped into three clusters, referred to as populations (P1, P2 and P3) (Fig. 1b, Supplementary Table 1 and Methods), each of which was characterized by the specific expression of hundreds of genes (Fig. 1c and Supplementary Table 2). Notably, only P2-specific genes enriched for adipogenesis-related functionality, including the adipogenic master regulator *Pparg* (Fig. 1c, Extended Data Fig. 1f, g and Supplementary Tables 2, 3). We further assessed the individual and combined expression of commonly used lineage marker genes (Supplementary Table 4 and Methods). We found that only *Fabp4*—which is associated with adipogenesis—was significantly differentially expressed across cells, showing high expression in P2 (Extended Data Fig. 1h and Supplementary Table 19). Consistently, P2 cells scored significantly higher on the adipogenic scale than did P1 or P3 cells (Extended Data Fig. 1i, Supplementary Table 19 and Methods). Virtually all cells scored high on the stem cell scale, as also exemplified by *Cd34* expression (Extended Data Fig. 1h, j and Supplementary Table 19), consistent with the selection of an ASPC-enriched cell pool. Finally, among genes previously used to mark pre-adipogenic populations[18], only *Pparg* and *Prdm16* expression was significantly correlated with *Fabp4* expression and enriched in P2 cells (Extended Data Fig. 1k, l and Supplementary Tables 5, 19).

To independently validate the three cell populations we detected, we assessed a larger number of Lin$^-$ SVF cells using the 10x Genomics Chromium platform. We considered 1,804 high-quality single cells (Fig. 1d, Extended Data Fig. 2a, b and Methods) and found that most of them reflected an ASPC state, on the basis of their high *Itgb1* (encoding CD29), *Cd34* and *Ly6a* (encoding SCA1) expression (Extended Data Fig. 2c). Consistently, mature adipocyte markers such as *Adipoq*, *Retn* and *Cidec* were virtually absent. We found that despite differences in cell stratification, capture and molecular assessment, these data largely recapitulated our initial findings (Fig. 1e, f, Supplementary Table 6 and Methods). The four clusters (G1–G4) that we detected could all be mapped back to the three previously described populations: G1 and G4 together approximately corresponded to P1, G2 to P2, and G3 to P3, with over 30% of the top 100 markers overlapping (Fig. 1f, Extended Data Fig. 2d, e, Supplementary Table 7 and Methods). This included the top-ranking transcription factors *Creb5* (in P1 and G1), *Peg3* (in P2 and G2) as well as *Meox2* (in P3 and G3). Consistently, G2-specific genes enriched for pathways related to adipogenesis (Supplementary Table 8), scored significantly higher on the adipogenic scale (Extended Data Fig. 2f and Supplementary Table 19) and included *Fabp4* (Fig. 1g, h and Extended Data Fig. 2g). *Pparg* was again strongly positively correlated with *Fabp4*, in contrast to most of the other (pre-)adipogenic markers with the notable exception of *Pdgfrb* (Fig. 1h, Extended Data Fig. 2g, h and Supplementary Tables 9, 19). *Cd34* and *Ly6a* were expressed at higher levels in G1 and G4 cells, and were significantly negatively correlated with *Pdgfrb*, *Fabp4* and *Pparg* expression (Fig. 1h and Supplementary Tables 9, 19). In summary, by using two distinct scRNA-seq-based approaches, we identified at least three subpopulations among Lin$^-$ SVF cells. The two major ones lie at opposite poles of stem cell-specific (*Cd34* and *Ly6a)* and pre-adipogenic (*Fabp4*, *Pparg* and *Cd36)* gene expression. These two subpopulations make up over 90% of the cells (Extended Data Fig. 2d) and can be further subclassified in increasingly homogenous cell groups, all of which are molecularly separated from a third and unrelated subpopulation (G3 or P3).

To assign adipogenic functionality to the three major subpopulations that we detected, we identified marker genes that encode surface proteins: *Cd55* and *Il13ra1* for P1 (and G1 and G4), *Aoc3* (encoding VAP1) and *Adam12* for P2 (and G2), and *F3* (encoding CD142) and *Abcg1* for P3 (and G3) (Fig. 2a and Extended Data Fig. 3a). Using FACS (Extended Data Fig. 3b), we isolated cellular fractions that were enriched for these respective markers and validated each subpopulation by quantitative PCR (Fig. 2b, Extended Data Fig. 3c and Supplementary Table 10). We then quantified lipid accumulation in response to a white fat differentiation cocktail for all ASPCs, marker-positive ASPCs and marker-negative ASPCs (Methods). We did not observe a difference in adipogenesis for P2 cells, whereas P1 cells showed an increased propensity and P3 cells a strongly decreased propensity to form adipocytes in

[1]Laboratory of Systems Biology and Genetics, Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL) and Swiss Institute of Bioinformatics, Lausanne, Switzerland. [2]Institute of Food Nutrition and Health, Eidgenössische Technische Hochschule Zürich, Schwerzenbach, Switzerland. [3]Swiss Stem Cell Foundation, Gentilino, Switzerland. [4]Concept-Clinic, Geneva, Switzerland. [5]These authors contributed equally: Petra C. Schwalie, Hua Dong, Magda Zachara. *e-mail: christian-wolfrum@ethz.ch; bart.deplancke@epfl.ch

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

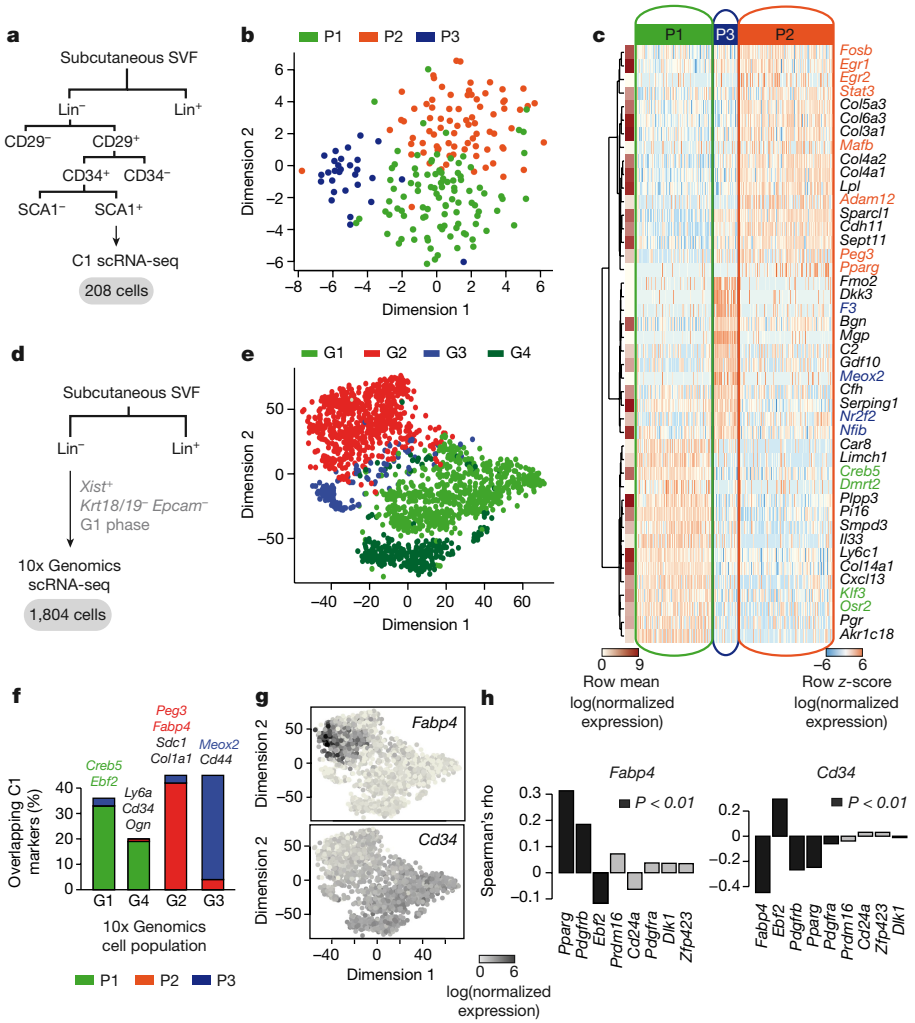


Fig. 1 | ScRNA-seq reveals the heterogeneity of subcutaneous ASPCs in the mouse. a, Schematic of FACS-based selection of ASPCs (Fluidigm C1). **b**, *t*-distributed stochastic neighbour embedding (*t*-SNE) 2D cell map (C1). $n = 208$ cells total, 3 biological replicates; populations P1, $n = 83$; P2, $n = 96$; P3, $n = 29$ cells. **c**, Heat map (blue-to-red) of expression of top 10 population markers and transcription factors among the top 100 population markers (in colour); hierarchically clustered rows, mean gene expression (white-to-red, left). $n = 208$ cells, 3 biological replicates. **d**, Schematic of FACS-based selection of Lin⁻ SVF cells (10x Genomics). **e**, *t*-SNE 2D cell map (10x Genomics); populations G1, $n = 699$; G2, $n = 664$; G3, $n = 122$; and G4, $n = 319$ cells. **f**, Percentage of the top 100 10x Genomics population markers that overlap with the top 100 C1 population markers; transcription factors and lineage markers are highlighted. **g**, *t*-SNE 2D cell map (10x Genomics) highlighting the expression of an adipogenic (*Fabp4*) and a stem cell (*Cd34*) marker (black). **h**, Correlation (Spearman's rho) between the expression of (pre-)adipogenic markers and *Fabp4* (left) or *Cd34* (right). In **e**–**h**, $n = 1{,}804$ cells, 1 biological replicate. See Extended Data Figs. 1, 2 and 8 for related results.

response to the same stimulus (Fig. 2c, d and Extended Data Fig. 3d, e). Consistently, we measured significantly higher levels of expression of adipogenic markers in P1 cells compared to ASPCs depleted of P1 cells, and the reverse was true for P3 cells (Extended Data Fig. 3f and Supplementary Tables 10, 19). None of the populations exhibited differences in nuclei number, suggesting that the adipogenic differences we observed were not due to distinct proliferative properties (Extended Data Fig. 3g and Supplementary Table 19).

CD142⁻ABCG1⁻ and CD142⁻ ASPCs (hereafter CD142⁻(ABCG1⁻) ASPCs) showed markedly higher differentiation compared to all ASPCs, which suggests that P3 cells inhibit adipogenesis (Fig. 2c, d and Extended Data Fig. 3d–f). To test this hypothesis, we performed a titration experiment in which CD142⁺ABCG1⁺ ASPCs were mixed with CD142⁻ABCG1⁻ ASPCs in ratios ranging from 0 to 100 per cent. We observed a nonlinear relation that showed a greater decrease in adipogenic differentiation than expected on the basis of a simple dilution effect (Fig. 2e, f). Again, we did not observe significant differences in nuclei numbers, which suggests that proliferation remained unaltered (Extended Data Fig. 4a and Supplementary Table 19). Thus, a fraction of ASPCs that is characterized by high CD142 and ABCG1 expression is refractory to adipogenesis and negatively regulates the adipogenic capacity of other ASPCs. A negatively modulatory function such as this is conceptually reminiscent of T regulatory cells or 'Tregs' that use their immunosuppressive capacity to maintain immune homeostasis and mediate peripheral tolerance[22]; therefore, we suggest that these 'adipogenesis-regulatory' cells be named 'Aregs'.

To characterize Aregs, we profiled the transcriptomes of all ASPCs, CD142⁺ ASPCs and CD142⁻ ASPCs at four time points: ex vivo, upon culture (after 5 and 24 h) and after adipogenic differentiation (Extended Data Fig. 4b–f, Supplementary Table 11 and Methods). This transcriptomic data further molecularly confirmed the non-adipogenic character of Aregs. Genes that have previously been associated[2] with adipogenesis—such as *Pparg*, *Fasn*, *Fabp4*, *Lpl* and *Fabp12*—were expressed at significantly lower levels in Aregs that were subjected to differentiation, when compared to both all ASPCs and to CD142⁻ ASPCs (Extended Data Fig. 4g–i and Supplementary Tables 11, 12, 19). The majority (64% of the differentially expressed genes) of Areg-specific expression patterns were maintained upon plating and only a further 20% of the differentially expressed genes changed their expression after culture for another day (Extended Data Fig. 4j). The genes that maintained their expression included those encoding the two surface markers used for the isolation of Aregs (*F3* and *Abcg1*),

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

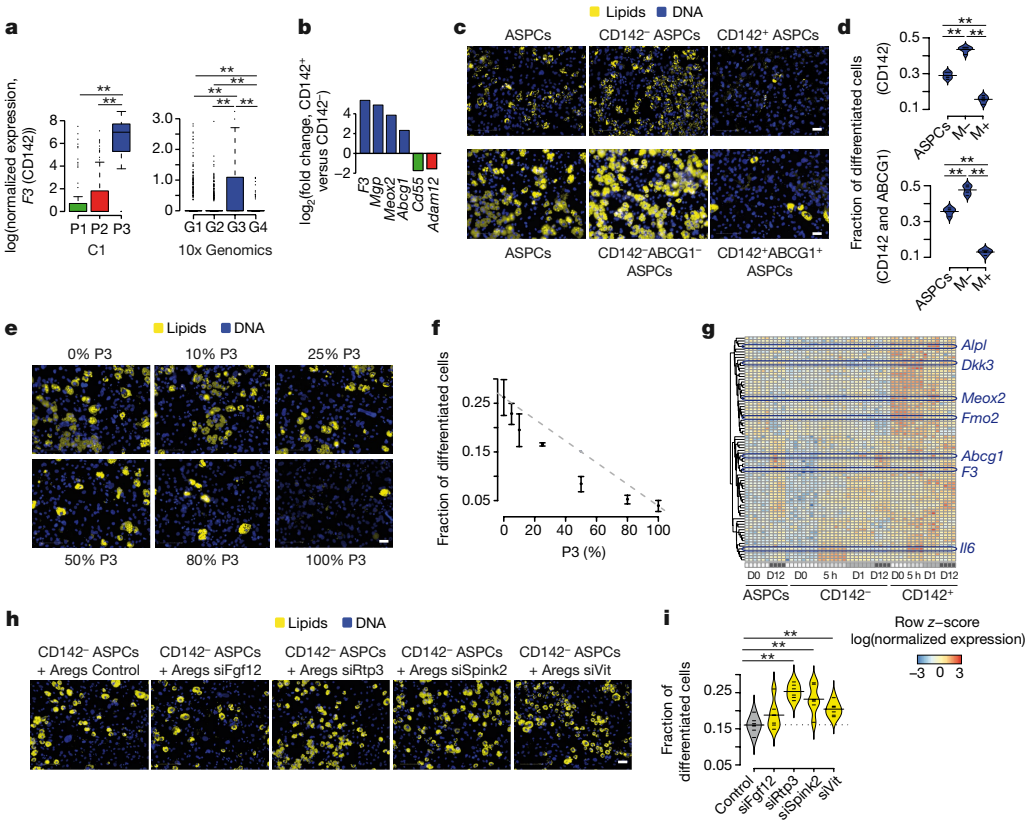


Fig. 2 | Aregs (CD142$^+$ABCG1$^+$ ASPCs) show a paracrine adipogenic inhibitory capacity. a, Expression of the P3 and G3 FACS marker *F3* (CD142) assessed using scRNA-seq; P1, $n = 83$ (green); P2, $n = 96$ (red); P3, $n = 29$ (blue) cells; G1, $n = 699$; G2, $n = 664$; G3, $n = 122$; and G4, $n = 319$ cells. $^{**}P \leq 0.01$, Wilcoxon rank-sum test. **b**, Fold-changes (CD142$^+$ versus CD142$^-$ ASPCs) of population-specific genes assessed using qPCR. **c**, Microscopy images of ASPC fractions after adipogenesis. **d**, Fraction of differentiated cells per ASPC type shown in **c**, $n = 4$–5 independent wells. M−, marker-negative, M+, marker-positive. In **c**, **d**, experiments were repeated at least three times, yielding similar results. **e**, Microscopy images of distinct ratios of Lin$^-$SCA1$^+$CD142$^+$ABCG1$^+$ (P3) and Lin$^-$SCA1$^+$CD142$^-$ABCG1$^-$ cells after adipogenesis. **f**, Mean and s.d. of the percentage of differentiated cells per titrated fraction shown in **e**; $n = 4$ independent wells. **g**, Heat map (blue-to-red) of expression of genes that show significantly higher expression levels in CD142$^+$ than in CD142$^-$ ASPCs after sorting (day 0, D0), plating (5 h) and culture (day 1, D1); expression levels upon adipogenesis (day 12, D12) are also included; lineage (osteogenic *Alpl* and inflammatory *Il6*) and genes among the top 10 scRNA-seq-derived markers (Fig. 1c) are highlighted; $n = 4$–8, 4 biological replicates, 1–3 independent wells each. **h**, Microscopy images after adipogenesis of Areg-depleted ASPCs co-cultured with Aregs with *Fgf12*, *Rtp3*, *Spink2*, *Vit* and control siRNA-mediated knockdown. **i**, Fraction of differentiated cells measured in the CD142$^-$ ASPCs underneath Aregs with the knockdowns shown in **h**. In **e**, **f**, **h**, **i**, experiments were repeated two times, yielding similar results. In **h**, **i**, $n = 6$, 2 biological replicates, 3 independent wells each. In all panels, nuclei are stained with DAPI (blue) and lipids are stained with LD540 (yellow). Scale bars, 50 μm (**d**, **i**). $^*P \leq 0.05$, $^{**}P \leq 0.01$, *t*-test. See Extended Data Figs. 3–5 for related results.

the early osteogenic marker *Alpl* and the inflammation-associated cytokine *Il6* (Fig. 2g). Forty-seven genes retained their specific expression across plating, culture and adipogenic differentiation, including those genes that encoded the transcription factor MEOX2, the secreted factor DKK3 and the enzyme FMO2, all among the top genes associated with P3 in the scRNA-seq data (Figs 1c, 2g). Consistent with the scRNA-seq results (Extended Data Fig. 1f, g), we found that genes that were expressed at high levels in Aregs immediately after sorting are characteristic of blood vessels and enrich for the 'Hedgehog signalling' pathway (Extended Data Fig. 4k–m and Supplementary Tables 3, 12). However, we did not observe consistently higher expression of endothelial marker genes in Aregs (Extended Data Fig. 4n).

We further investigated the mechanism by which Aregs exert their adipogenesis-suppressive effect by testing whether the inhibition requires direct cell–cell contact. Transwell experiments in which ASPCs, Aregs or Areg-depleted ASPCs were inserted above ASPCs revealed that adipogenesis differences were maintained in the absence of direct contact (Extended Data Fig. 5a, b), which suggests a paracrine signalling mechanism. To identify genes that may mediate this signal, we performed a loss-of-function screen targeting 23 genes that showed high expression in Aregs compared to Areg-depleted ASPCs and were expressed only at low levels in mature adipocytes (Supplementary Table 13 and Methods). We subjected plated SVF cells to short interfering RNA (siRNA)-mediated reduction in expression of each of these genes, reasoning that knockdown would be effective only in Aregs. Thus, any changes in differentiation would be attributable to alterations in Areg activity. We found that the knockdown of *Rtp3*, *Spink2*, *Fgf12* and *Vit* significantly increased the differentiation of plated SVF cells, but not cell number (Extended Data Fig. 5c–f and Supplementary Table 19), which suggests that these genes modulate the inhibitory potential of Aregs. To validate these findings, we used transwells comparing the anti-adipogenic effect of wild-type Aregs to the anti-adipogenic effect in Aregs in which each of these genes was knocked down. Loss of *Rtp3* and, to a lesser extent, *Spink2* and *Vit* lowered the paracrine inhibitory potential of Aregs (Fig. 2h, i and Extended Data Fig. 5g–j). Our findings therefore support a paracrine mechanism of action, in which these genes may act upstream of one or several factors secreted by Aregs or, in the case of *Spink2*, as a direct signalling molecule.

To investigate whether Aregs constitute a conserved cell population that functions similarly in humans, we used a FACS isolation strategy based on CD142 that was analogous to the strategy we used for the mouse model (Extended Data Fig. 6a and Methods). Across three individuals, CD142$^+$ cells exhibited higher levels of *F3* gene expression when compared to CD142$^-$ cells (Extended Data Fig. 6b and

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

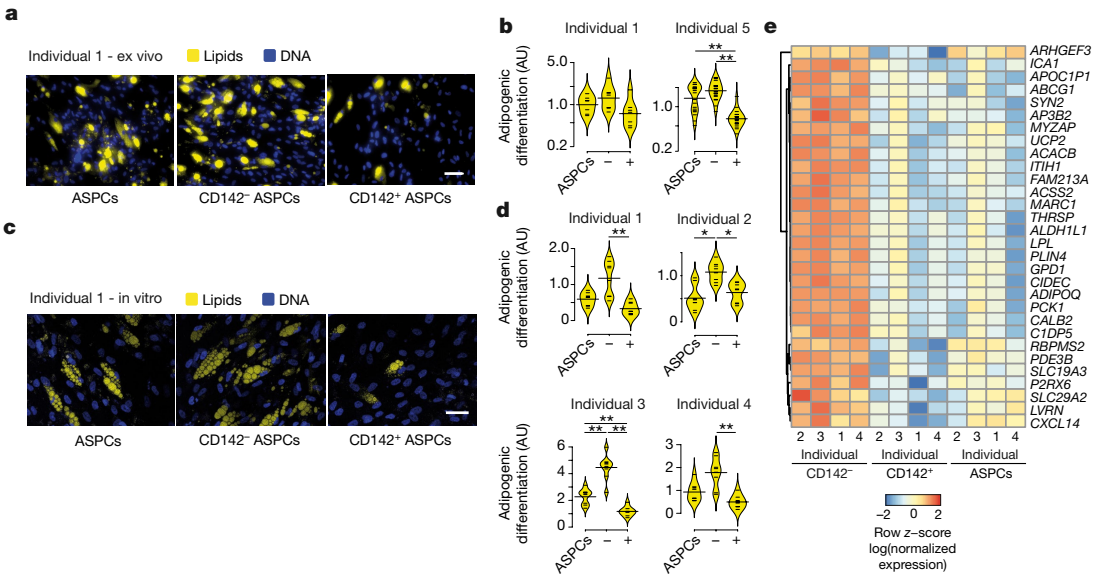


Fig. 3 | Aregs and their inhibitory capacity are conserved in humans. **a**, Microscopy images of ex vivo human ASPCs, CD142$^-$ ASPCs and CD142$^+$ ASPCs after adipogenesis. **b**, Adipogenic differentiation (arbitrary units, AU) per each fraction shown in **a** (individual 1, $n = 7$ fields of view) and another individual (individual 5, $n = 20$ fields of view). −, CD142$^-$; +, CD142$^+$. **c**, Microscopy images of in vitro human (individual 1) ASPCs, CD142$^-$ ASPCs and CD142$^+$ ASPCs after adipogenesis. In **a**, **c**, experiments were repeated at least three times (distinct individuals), yielding similar results. **d**, Adipogenic differentiation per each fraction shown in **c**, and in three other individuals (individuals 1 to 4, $n = 8$ fields of view). −, CD142$^-$; +, CD142$^+$. **e**, Expression of genes that were most-highly expressed in CD142$^-$ versus CD142$^+$ ASPCs upon adipogenesis (day 12, D12); $n = 4$ biological replicates (distinct individuals). In all panels, nuclei are stained with Hoechst (blue) and lipids are stained with Bodipy (yellow). Scale bars, 50 μm, $*P \leq 0.05$, $**P \leq 0.01$, Wilcoxon rank-sum test. See Extended Data Fig. 6 for related results.

Supplementary Tables 14, 19), validating our approach. We observed only a mild enrichment of *ABCG1* expression (Extended Data Fig. 6b), which may point to species-specific gene expression differences. Similar to mouse CD142$^+$ ASPCs, human CD142$^+$ ASPCs were refractory to adipogenic differentiation, as assessed by lipid accumulation (Fig. 3a, b and Methods) and marker gene expression (Extended Data Fig. 6b and Supplementary Table 14). Consistent with the mouse data, *F3* expression increased during adipogenesis, which negated the difference in expression between the CD142$^+$ and CD142$^-$ populations (Extended Data Fig. 6b). We further examined whether the adipogenic inhibitory property is maintained after culture by separating the CD142$^+$ population from the CD142$^-$ population in cultured (passage 1) ASPCs (Extended Data Fig. 6c), and then inducing adipogenesis. Population differences were consistent across four individuals (Fig. 3c, d). We observed significantly larger numbers of mature adipocytes among CD142$^-$ cells than among all ASPCs, which suggests that human CD142$^+$ ASPCs also have an inhibitory capacity (Fig. 3c, d and Supplementary Table 19). Transcriptomic profiling of induced cells revealed hundreds of differentially expressed genes among the three fractions, with CD142$^-$ ASPCs exhibiting the highest number of differences (Extended Data Fig. 6d and Supplementary Table 15). The adipocyte-like phenotype of CD142$^-$ cells was globally confirmed by the high expression of genes associated with adipogenic functionality (Fig. 3e, Extended Data Fig. 6e–i and Supplementary Table 16), including the adipogenic markers *ADIPOQ*, *CIDEC* and *LPL*. Finally, we performed a cross-species experiment in which human Lin$^-$ SVF cells were co-cultured with mouse Aregs or Areg-depleted ASPCs. We observed a reduction in adipogenesis in the transwells containing Aregs (Extended Data Fig. 6j, k), which suggests that the paracrine signal is conserved across species. Taken together, our data suggest that Aregs are functionally conserved in humans.

Having characterized Aregs based on cell-culture experiments, we next aimed to study their distribution and effect in vivo. Immunofluorescence staining of cryosections of the subcutaneous adipose tissue of mice revealed a CD142 signal that indicated a perivascular sheath, which featured small individual cells and was located within the adipose parenchyma but not the lymph nodes (Fig. 4a and Extended Data Fig. 7a–d, arrows). By combining staining for CD142 with staining for the stem cell marker SCA1, we identified a small fraction of CD142$^+$SCA1$^+$ cells that are also located perivascularly (Fig. 4b–d and Extended Data Fig. 7e–g, arrows). This is consistent with our gene expression data, which suggest that Aregs constitute a small fraction of all ASPCs as well as a link to blood vessels (Extended Data Figs 1g, 2d and 4l). We therefore hypothesize that the CD142$^+$SCA1$^+$ cells we detected correspond to Aregs.

Using a FACS strategy analogous to the one used above, we found that Aregs were also present in visceral fat depots (Fig. 4e and Methods). The visceral SVF contained a significantly higher proportion of CD142$^+$ABCG1$^+$ cells, when compared to the subcutaneous SVF (Fig. 4e and Supplementary Table 19). Furthermore, we found that obese *ob/ob* mice have significantly more Aregs than do lean mice, in both the subcutaneous and the visceral adipose depots (Fig. 4e and Supplementary Table 19). To assess whether the presence of Aregs exerts an anti-adipogenic effect in vivo, we embedded (1) total SVF or CD142$^-$ABCG1$^-$ SVF cells and (2) Lin$^-$SCA1$^+$ or CD142$^-$Lin$^-$SCA1$^+$ SVF cells in Matrigel, injecting them into the left and right subcutaneous layer of the mouse abdomen (Methods). After implantation, the pad-bearing mice were fed a high-fat diet for three weeks to induce adipogenesis. Consistent with our in vitro data, we observed a notable difference between the Matrigel pads for all mice (Fig. 4f–h and Extended Data Fig. 7h–j). The absence of Aregs either from total SVF cells or from ASPCs led to a significantly higher number of mature adipocytes being formed in the implants, as also revealed by the much whiter appearance of Matrigel pads containing CD142$^-$(ABCG1$^-$) cells (Supplementary Table 19). The staining of implant cryosections for endothelial cells using isolectin IB4 revealed a similar degree of vascularisation (Extended Data Fig. 7k, l). These findings strongly suggest that Aregs control the formation of mature adipocytes in vivo.

Our single-cell transcriptomics-based approach using a mouse model revealed at least three cell populations in the subcutaneous ASPC-enriched SVF. The largest one was CD55$^+$, and showed enrichment of the adipogenic stem cell markers *Cd34* and *Ly6a* as well as a high adipogenic differentiation capacity in vitro. A recent study has

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

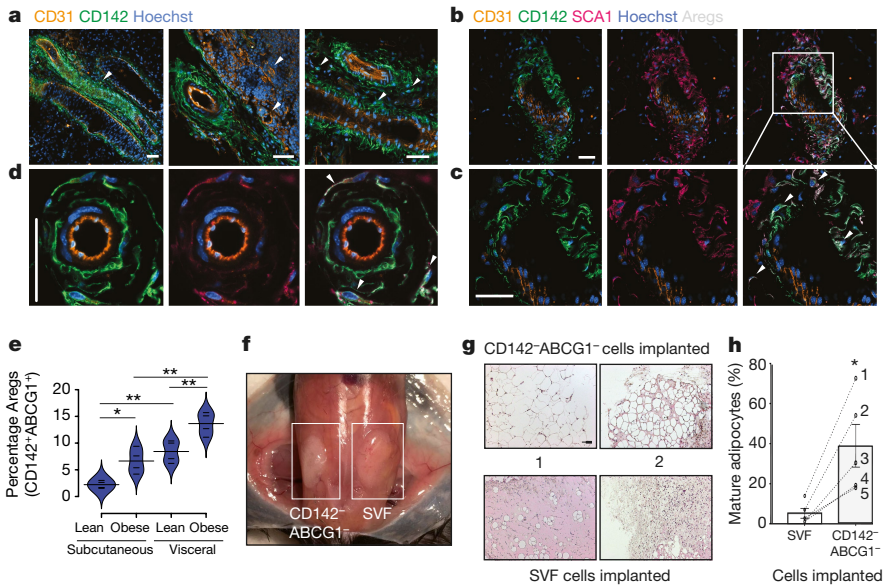


Fig. 4 | Aregs locate proximal to the vasculature and inhibit adipogenesis in vivo. a–**d**, Microscopy images of the in vivo localization of CD31 (orange), CD142 (green) and SCA1 (pink) in subcutaneous white adipose tissue in mice. Nuclei are stained with Hoechst (blue). Scale bars, 50 μm. Experiments were repeated at least three times, yielding similar results. Arrows in **a** indicate $CD142^+$ around big vessels within the adipose parenchyma (left), $CD142^+$ absence in the lymph nodes (middle) and $CD142^+$ in individual cells (right); arrows in **c, d** indicate individual cells presenting both CD142 and SCA1 staining (white). **e**, Fraction of Aregs ($CD142^+ABCG1^+$ ASPCs) in the subcutaneous and visceral SVF from lean and obese mice; $n = 4$ biological replicates. **f**–**h**, Total SVF and Areg-depleted ($CD142^-ABCG1^-$) SVF cells were implanted into the subcutaneous adipose depot in Matrigel. **f**, Image of implantation sites after three weeks of high-fat diet. **g**, Microscopy images of Matrigel plugs with total SVF or $CD142^-ABCG1^-$ SVF cells. Scale bar, 100 μm. **h**, Mean and s.d. of the percentage of mature adipocytes that developed within the plugs ($n = 5$ biological replicates) analysed using CellProfiler[25]. $^*P \leq 0.05$, $^{**}P \leq 0.01$ paired *t*-test. See Extended Data Fig. 7 for related results.

identified a homologous subpopulation in the human SVF, which exhibited an enhanced regenerative potential and was dysfunctional in diabetes[23]. These $CD55^+$ cells may therefore exhibit the highest plasticity of all Lin^- SVF cells in both mice and humans, a hypothesis that warrants further investigation. We determined that a small (less than 10% of Lin^- SVF cells) $CD142^+$ subpopulation suppresses adipocyte formation in vitro and in vivo. To our knowledge, we provide the first molecular characterization and functional follow-up of this $CD142^+$ subpopulation and show that the mechanism of action of this subpopulation possibly involves the genes *Spink2*, *Rtp3*, *Vit* and/or *Fgf12*. Our findings suggest that depot- and condition-specific differences in hyperplasia might not be solely attributable to alterations in the number of ASPCs, but instead might also be influenced by the number of Aregs, as these latter cells may control the de novo adipogenic capacity of adipose tissue. As adipocytes can also arise in tissues such as bone marrow and muscle[24], it is tempting to speculate that de novo adipogenesis in these tissues is also controlled through the presence of Areg-like cells. Together, our findings point to a potentially critical role for Aregs in modulating the plasticity and metabolic signature of distinct fat-cell-containing systems, within which they may constitute essential components of the elusive adipogenic precursor niche.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0226-8.

Received: 3 May 2017; Accepted: 16 May 2018;
Published online 20 June 2018.

1. Müller, S., Kulenkampff, E. & Wolfrum, C. in *Metabolic Control* (ed. Herzig, S.) 251–263 (Springer International Publishing, Cham, 2015).
2. Rosen, E. D. & Spiegelman, B. M. What we talk about when we talk about fat. *Cell* **156**, 20–44 (2014).
3. Crisan, M. et al. A perivascular origin for mesenchymal stem cells in multiple human organs. *Cell Stem Cell* **3**, 301–313 (2008).
4. Tang, W. et al. White fat progenitor cells reside in the adipose vasculature. *Science* **322**, 583–586 (2008).
5. Vishvanath, L. et al. $Pdgfr\beta^+$ mural preadipocytes contribute to adipocyte hyperplasia induced by high-fat-diet feeding and prolonged cold exposure in adult mice. *Cell Metab.* **23**, 350–359 (2016).
6. Zannettino, A. C. W. et al. Multipotential human adipose-derived stromal stem cells exhibit a perivascular phenotype *in vitro* and *in vivo*. *J. Cell. Physiol.* **214**, 413–421 (2008).
7. Cai, X., Lin, Y., Hauschka, P. V. & Grottkau, B. E. Adipose stem cells originate from perivascular cells. *Biol. Cell* **103**, 435–447 (2011).
8. Gupta, R. K. et al. *Zfp423* expression identifies committed preadipocytes and localizes to adipose endothelial and perivascular cells. *Cell Metab.* **15**, 230–239 (2012).
9. Jiang, Y., Berry, D. C., Tang, W. & Graff, J. M. Independent stem cell lineages regulate adipose organogenesis and adipose homeostasis. *Cell Reports* **9**, 1007–1022 (2014).
10. Tran, K.-V. et al. The vascular endothelium of the adipose tissue gives rise to both white and brown fat cells. *Cell Metab.* **15**, 222–229 (2012).
11. Majka, S. M. et al. De novo generation of white adipocytes from the myeloid lineage via mesenchymal intermediates is age, adipose depot, and gender specific. *Proc. Natl Acad. Sci. USA* **107**, 14781–14786 (2010).
12. Billon, N. et al. The generation of adipocytes by the neural crest. *Development* **134**, 2283–2292 (2007).
13. Sowa, Y. et al. Adipose stromal cells contain phenotypically distinct adipogenic progenitors derived from neural crest. *PLoS ONE* **8**, e84206 (2013).
14. Su, X. et al. Fascia origin of adipose cells. *Stem Cells* **34**, 1407–1419 (2016).
15. Berry, R., Jeffery, E. & Rodeheffer, M. S. Weighing in on adipocyte precursors. *Cell Metab.* **19**, 8–20 (2014).
16. Sanchez-Gurmaches, J. & Guertin, D. A. Adipocyte lineages: tracing back the origins of fat. *Biochim. Biophys. Acta* **1842**, 340–351 (2014).
17. Cristancho, A. G. & Lazar, M. A. Forming functional fat: a growing understanding of adipocyte differentiation. *Nat. Rev. Mol. Cell Biol.* **12**, 722–734 (2011).
18. Jeffery, E. et al. Characterization of Cre recombinase models for the study of adipose tissue. *Adipocyte* **3**, 206–211 (2014).
19. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620 (2015).
20. Rodeheffer, M. S., Birsoy, K. & Friedman, J. M. Identification of white adipocyte progenitor cells *in vivo*. *Cell* **135**, 240–249 (2008).
21. Hudak, C. S. et al. Pref-1 marks very early mesenchymal precursors required for adipose tissue development and expansion. *Cell Reports* **8**, 678–687 (2014).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

22. Vignali, D. A. A., Collison, L. W. & Workman, C. J. How regulatory T cells work. *Nat. Rev. Immunol.* **8**, 523–532 (2008).
23. Rennert, R. C. et al. Microfluidic single-cell transcriptional analysis rationally identifies novel surface marker profiles to enhance cell-based therapies. *Nat. Commun.* **7**, 11945 (2016).
24. Shook, B. et al. The role of adipocytes in tissue regeneration and stem cell niches. *Annu. Rev. Cell Dev. Biol.* **32**, 609–631 (2016).
25. Gubelmann, C. et al. Identification of the transcription factor ZEB1 as a central component of the adipogenic gene regulatory network. *eLife* **3**, e03346 (2014).

Acknowledgements We thank J. Auwerx, W. Chen, E. Dorcey, N. Gheldof, O. Naveiras and K. Schoonjans for constructive discussions and carefully reading the manuscript. This research was supported by the Human Frontier Science Program LT001032/2013 (to P.C.S.), the Swiss National Science Foundation Grants (#31003A_162735 and 31003A_162887), the Kristian Gerhard Jebsen Foundation for Metabolic Research (to B.D.) and by institutional support from the Swiss Federal Institute of Technology in Lausanne (EPFL) and Zürich (ETH). We thank the EPFL FCCF, BIOP (especially R. Guiet and O. Burri), HCF and GECF for cell sorting, imaging, histology and sequencing support, as well as the VITAL-IT platform (University of Lausanne) for computational support.

Reviewer information *Nature* thanks O. Stegle and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions B.D., C.W., P.C.S., M.Z. and H.D. designed the study and wrote the manuscript. P.C.S. conducted all analyses related to transcriptomics. N.A., J.R., H.D. and M.Z. performed the single-cell experiments. H.D. and M.Z. conducted all FACS, qPCR, cell culture and related imaging assays. H.D. and W.S. performed all experiments related to transplants. H.D. performed all siRNA-based knockdown experiments. M.Z. and J.R. conducted all histological assays. C.C., K.-U.S. and G.S. provided access to human samples and helped with processing them. M.Z. and D.A. performed all experiments related to human cells. All authors read and approved the final manuscript.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0226-8.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0226-8.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to C.W. or B.D.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment, except for the quantification of in vivo implantation experiments (Fig. 4f–h and Extended Data Fig. 7h–l).

Bioethics. All mouse experiments were conducted in strict accordance with the Swiss law and all experiments were approved by the ethics commission of the state veterinary office (VD2984/2015, 60/2012, 43/2011, 80/2014). The work on human ASPC cultures derived from human lipoaspirate samples was approved by the ethical commission of Canton Ticino (CE 2961 from 22.10.2015) and conforms to the guidelines of the 2000 Helsinki declaration. The anonymized samples were collected under signed informed consent.

Generation of the DLK1 (also known as PREF1)–RFP mouse. The new mouse strain Tg(Pref1-CreER)426 Biat (Pref1-CreER) was generated using the bacterial artificial chromosome (BAC) RP24-334E14 (BACPAC Resources Center) according to previously published methods[26,27].

Isolation of the mouse SVF. Subcutaneous and visceral adipose tissue depots were dissected from male and female 8–11-week-old wild-type, DLK1–RFP or male *ob/ob* C57BL/6J mice into ice-cold PBS. The tissue was finely minced using scissors, transferred into collagenase (Sigma-Aldrich #C6885-1G, 2 mg/ml of collagenase buffer (25 mM $NaHCO_3$, 12 mM KH_2PO_4, 1.2 m $MgSO_4$, 4.8 mM KCl, 120 mM NaCl, 1.4 mM $CaCl_2$, 5 mM Glucose, 2.5% BSA, pH = 7.4)) and incubated for 1 h at 37 °C under agitation. Following a 10-min centrifugation at 400 *g* at room temperature, floating mature lipid-filled adipocytes were aspirated and the cell suspension was filtered through a 40-μm cell strainer to ensure a single cell preparation. Next, a red blood cell lysis was performed by incubating the pelleted cells in the red blood cell lysis buffer (154 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM EDTA) for 5 min, followed by two washes (a 5-min centrifugation at 400*g*, room temperature) with FACS buffer (PBS with 3% fetal bovine serum (FBS) (Gibco #10270-106), 1 mM EDTA, 1% penicillin–streptavidin (Gibco #15140122)).

Isolation of human SVF. Fresh lipoaspirates were washed twice with DPBS with calcium and magnesium (Gibco #14040091) in 100-ml syringes (VWR International #720-2528) (50 ml lipoaspirate per 40 ml DPBS with calcium and magnesium) and incubated with 0.28 U/ml of liberase TM (Roche #05401119001) for 45 min at 37 °C under agitation. The digested tissue was mixed with 1% human albumin (CSL Behring) in DPBS −/− (Gibco #14190094) (50 ml lipoaspirate per 40 ml 1% human albumin in DPBS −/−) and shaken vigorously to liberate the stromal cells (procedure performed twice). The aqueous phase was recovered and centrifuged (all centrifugations performed at 400 *g* for 5 min at room temperature). The cell pellet was resuspended in 15 ml remaining buffer and filtered through a 100-μm and then a 40-μm cell strainer to ensure a single-cell preparation, centrifuged and resuspended in 5 ml of 5% human albumin (CSL Behring). The viability and the number of nucleated cells in the cell suspension obtained was determined using a Nucleostainer, after which a red blood cell lysis was performed using VersaLyse solution (Beckman Coulter #A09777) according to the manufacturer's recommendations.

scRNA-seq. For the initial (Fluidigm C1) scRNA-seq experiments, we used the DLK1–RFP mouse strain, as recent results have suggested that adipogenic progenitors may be marked by the expression of *Dlk1*[21]. Specifically, in this strain, RFP is induced in *Dlk1*-expressing cells upon feeding with tamoxifen. Thus, all cells in which the *Dlk1* promoter has been active become RFP^+ and can be easily isolated by FACS. We investigated both RFP^+ and RFP^- cells to explore the differences (if any) in cell composition and molecular characteristics between the two cell populations. However, as no differences were observed, we treated and discussed the data irrespective of RFP status. Cells (550 RFP^+ and 550 RFP^-, 1,100 in total) at a concentration of 200 cells/μl were sorted by FACS directly into the C1 Single-Cell Auto Prep Array IFC 10–17 μm chip (Fluidigm #100-5760, C1) for 96 cells with the efficiency of ~10%. The presence of the cells was validated by microscopy visualization (Olympus Cell Xcellence, objective: UPLFLN 10×/0.30). Single-cell cDNA synthesis was performed on the C1 Single-Cell Auto Prep system (Fluidigm) according to the manufacturer's specifications. Single-cell libraries were multiplexed and sequenced across 3 lanes of HiSeq 2000 (Illumina) using 100-bp single-end sequencing. All results relating to the Fluidigm C1 single-cell experiment (Figs 1a–c, 2a and Extended Data Figs 1, 2a and 8) are based on 3 biological replicates performed on 3 different days, each stemming from 3–6 male and female 10–11-week-old mice.

A second scRNA-seq experiment was performed using the Chromium Single Cell Gene Expression Solution (10x Genomics), following the manufacturer's protocol. The SVF was isolated from six 8-week-old mice (3 males and 3 females) as described above. Cells were stained with the anti-mouse antibodies CD31-AF488, CD45-AF488 and TER119 AF488 (Biolegend #303110, #304017 and #116215, respectively), and 280,000 Lin^- cells were isolated using a Becton Dickinson FACS Aria III sorter. Cells were washed and resuspended in 250 μl FACS buffer (PBS, 2% FBS, 1 mM EDTA), targeting the required 1,000 cells/μl concentration, accounting for a 10–20% loss. We pipetted 9.7 μl cell suspension (concentration of 913 cells/μl, ~8,800 cells), targeting the recovery of ~5,000 cells. Single-cell RNA-seq libraries were obtained following the 10x Genomics recommended protocol, using the reagents included in the Chromium Single Cell 3′ v2 Reagent Kit. Libraries were sequenced on the NextSeq 500 v2 (Illumina) instrument using 150 cycles (18 bp barcode + UMI, and 132-bp transcript 3′ end), obtaining ~5×10^8 raw reads.

scRNA-seq data analysis. For the C1 scRNA-seq experiments, FASTQ files containing 100-bp-long single-end sequenced tags (reads) from 3 replicates of 96 cells each were trimmed and filtered using prinseq 0.20.3[28] with the parameters '-custom_params 'A 70%;T 70%;G 70%;C 70%' -trim_tail_left 36 -trim_tail_right 36 -lc_method dust -lc_threshold 45 -min_gc 1 -out_format 3' and cutadapt 1.5[29] with the parameters '-m 36 -q 20' and the Nextera adaptor sequence. The retained tags were evaluated using FastQC v.0.11.2 and aligned to the Ensembl 84[30] gene annotation of the NCBI38/mm10 mouse genome using STAR 2.4.0g[31] with the parameters '–runThreadN 4 –runMode alignReads –outFilterType BySJout –outFilterMultimapNmax 20 –alignSJoverhangMin 8 –alignSJDB overhangMin 1 –outFilterMismatchNmax 999 –outFilterMismatchNoverLmax 0.04 –alignIntronMin 20 –alignIntronMax 1000000 –alignMatesGapMax 1000000 g –outSAMtype BAM SortedByCoordinate'. The number of tags per gene was calculated using htseq-count 0.6.0[32] with the parameters 'htseq-count -m intersection-nonempty -s no -a 10 -t exon -i gene_id'. Cells that appeared as doublets in the microscopy images (33) and those that had <40% or <400,000 aligned reads were excluded, which resulted in a total of 208 cells, of which 63 were RFP^- according to the microscopy images. Genomic alignment rates, number of detected genes per cell, RNA spike-in recovery as well as correlations between replicate experiments and those with bulk population (control) samples suggested that our data were of high quality (Extended Data Fig. 1a–d). RFP^+ cells showed significantly higher expression of *RFP* than RFP^- cells ($P < 0.01$, Wilcoxon rank-sum test, data not shown). For each gene, expression estimates per gene were expressed as log-transformed counts per million (cpm) (log_cpm/log(norm_expr) in Fig. 1c and Extended Data Fig. 1), by dividing total tags per gene by the total number of gene-aligned reads per cell and taking the $\log(x+1)$ of this value. Using the marker genes listed in Supplementary Table 4, we calculated 'scores'—a single numeric value representative of the expression of multiple marker genes—as the sum of log_cpms across all markers in a category. We further filtered and normalized the data using the package M3Drop_1.0.1[33] and the function M3DropCleanData() with the parameters 'is.counts=T, min_detected_genes=3000', obtaining 17,287 genes and 208 cells. We used the function M3DropDropoutModels from the M3Drop package to fit the modified Michaelis–Menten equation, a logistic regression (logistic) or a double exponential (ZIFA) function to the relationship between mean expression and dropout-rate (proportion of zero values). After visual inspection of the three fits and examination of the sum of squared residuals and sum of absolute residuals, we determined that the Michaelis–Menten method produced the best fit to our data. We therefore used it to estimate genes that had a significantly higher number of drop-outs than expected by chance (referred to as differentially expressed genes above, and informative genes below) using the M3Drop function M3DropDifferentialExpression() at a false discovery rate (FDR) of 0.05, which resulted in 527 genes (listed in Supplementary Table 1). We used these genes to obtain a 2D representation of the cells, while maintaining the similarity relationships between them using *t*-SNE[34] as implemented in the package Rt-SNE_0.11[34]. Cluster analysis, including silhouette analysis, was performed using the SC3_1.3.6 package[35] on the 527 differentially expressed genes, using $k = 3$ or 4 (Fig. 1b and Extended Data Fig. 8a). P1 and P3 were stable across different cluster number choices, whereas P2 could be further subdivided. The cell grouping we obtained did not correspond to biological replicates, which suggests that the main signal captured by the selected genes is biological and not technical (Extended Data Fig. 8b, c). Silhouette analysis supported a three-cluster partitioning (Extended Data Fig. 8d), which also corresponded closely to the distribution of *Fabp4* expression. Marker genes per cluster were obtained using M3DropGetMarkers() function, with the top 200 included in Supplementary Table 2 and used for functional enrichment analysis (Supplementary Table 3). To test the robustness of our results to methodological choices, we alternatively used a set of 1,827 biologically highly variable (>0.5, FDR 0.05) genes calculated using the scran_1.2.2 package[36] instead of the drop-out-based gene selection. We found that 520 of these genes were among the 527 significantly differentially expressed genes described above (Extended Data Fig. 8e). The clustering was highly similar to that previously described, with P3 being largely unchanged and ~20% of P1 cells attributed as P2 cells (Extended Data Fig. 8f). Over 60% of the top 100 markers identified for each of the three populations were identical (Extended Data Fig. 8g) and the *t*-SNE projection placed P3 outside the P1–P2 space (Extended Data Fig. 8h). In Fig. 1c, we displayed in colour the top 10 markers per cluster, additional transcription-factor coding genes that were among the top 100 markers per cluster and highlighted genes that code for transcription factors and cell surface proteins. Heat maps were generated using gplots_3.0.1 and the function heatmap.2() with the parameters 'scale="row",

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Colv = F, Rowv = as.dendrogram (cluster)' (per row *z*-score transformed log(normalized expression), blue-to-red). Row means per gene are displayed on the left, in white-to-dark-red. All correlations were calculated based on log_cpm values, with the function cor() and the parameters 'method="Spearman'. Multiple testing correction using the function 'p.adjust and the parameters 'method = Bonferroni') was applied for *Fabp4* and *Cd34* expression correlations, respectively.

The 10x Genomics scRNA-seq data was processed using cellranger-2.1.0, default parameters and the mouse NCBI38/mm10 genome. Molecular counts were obtained for 2,919 cells (filtered matrix), with 174,362 mean reads/cell, an average of 60.5% reads mapping to the transcriptome and 3,404 median genes detected per cell. We also filtered outlier cells using the median absolute deviation from the median total library size (logarithmic scale) as well as total gene numbers (logarithmic scale), as implemented in scran[36], using a cutoff of 3 (isOutlier, nmads = 3). log(normalized expression) values were obtained using size factors per cell, estimated with scran. The majority of the cells expressed high levels of the housekeeping gene *Actb*, as well as the ASPC marker genes *Cd34*, *Ly6a* and *Itgb1* (Extended Data Fig. 8i). An initial analysis using all cells revealed that one of the major sources of variation was related to the expression of the gene *Xist*, as exemplified by the strong clustering of cells expressing *Xist* in the 2D *t*-SNE projection (Extended Data Fig. 8j). As *Xist* was expressed (>2 reads/cell) in 93% of the C1 cells, for comparative consistency we included only cells that expressed *Xist* (≥2 molecular counts per cell) in the reported results. We also noted a strong cell division signal in a small (27) number of cells, as quantified by Cyclone[37], and exemplified by *Mki67* expression (Extended Data Fig. 8j). As the C1 data contained cells that were predicted only to be in G1 phase (with a single exception), we also excluded the cells predicted to be in S or G2–M phase. Finally, as we noticed high levels of expression of epithelial-specific genes in a small subset of cells (again, dissimilar to the C1 experiment), we excluded all cells showing any (>0) *Epcam* and *Krt19* or *Krt18* expression. The final reported dataset consists of 1,804 cells (Fig. 1d). Genomic alignment rates and number of detected genes/cell suggested that our data were of high quality (Extended Data Fig. 2a, b). We further used both methods detailed above (M3Drop and highly variable genes) to determine a set of informative genes, and obtained a set of 550 genes with a significantly higher number of drop-outs than expected by chance (FDR = 0.0001, a stringent cutoff designed to exclude the selection of many genes expressed at very low levels, given that the 10x Genomics data showed a different drop-out distribution as compared to the C1 data) and 223 highly variable genes (FDR 0.1 and biological variability >0.3). Combining these two sets of genes gave 631 unique informative genes (Supplementary Table 6), which we used to perform consensus clustering using SC3_1.3.6 and to display the cells in 2D using *t*-SNE, similar to as described above. Although the sum of the squared differences decreased rapidly from using two to using three or four clusters (Extended Data Fig. 8k), silhouette analysis suggested a local maximum at four clusters (Extended Data Fig. 8l). We therefore used $k = 4$ in the manuscript. We note that a larger choice of k could provide further insight into additional subgroupings of the reported populations. Marker genes per cluster were again obtained using M3DropGetMarkers() function; the top 200 are included in Supplementary Table 7, and were used for functional enrichment analysis. The top 100 marker genes were used to compare the clusters (groups, G) with the previous ones (populations, P) (Fig. 1f and Extended Data Figs 2e, 8m), revealing >30% overlap for each population with its respective group.

FACS-based cell isolation of mouse cells. The isolated single-cell suspension was diluted to 0.75 or 1×10^7 cells/ml with FACS buffer (PBS with 3% FBS, 1 mM EDTA, 1% penicillin–streptavidin) and the following fluorophore-conjugated antibodies were added (in titration-determined quantities, Supplementary Table 17): anti-mouse CD31-AF488, anti-mouse CD45-AF488, anti-mouse TER119-AF488 (BioLegend #303110, #304017 and #116215, respectively) for selecting the Lin^- population; anti-mouse SCA1-PE-Cy7 (BioLegend #122513), anti-mouse CD34-BV421 (BioLegend #119321) and anti-mouse CD29-PerCP-efluor710 (eBiosciences #46-0291) to enrich the Lin^- population with ASPCs; anti-mouse CD55-APC (BioLegend #131802), anti-mouse VAP1 (Abcam #ab81673) conjugated with allophycocyanin (APC) (Lightning-Link Allophycocyanin (APC) Conjugation Kit (Innova Biosciences #705-0030), anti-mouse CD142-PE (SinoBiological #50413-R001) and anti-ABCG1 antibody (Invitrogen #PA5-13462) conjugated with APC for separating populations negative and positive for the given marker. The cells were incubated with the cocktail of antibodies on ice for 20 min protected from light, after which they were washed and stained with DAPI (Sigma #D9542) or propidium iodide (Molecular Probes #P3566) for assessing viability, and subjected to FACS using a Becton Dickinson FACSAria II sorter. Compensation measurements were performed for single stains using compensation beads (eBiosciences #01-2222-42).

The following gating strategy was applied while sorting the cells: first, the cells were selected based on their size and granulosity or complexity (side and forward scatter), and then any events that could represent more than one cell were eliminated. Next, the Lin^-($CD31^-CD45^-TER119^-$) population was selected, followed by $Lin^-SCA1^+CD34^+CD29^+$ (C1), $Lin^-SCA1^+CD34^+$, Lin^-SCA1^+, or Lin^- (10x) selection, which were used as controls for the further analysed populations that were negative or positive for a given marker: $Lin^-SCA1^+CD55^-$ ($CD55^-$) and $Lin^-SCA1^+CD55^+$ ($CD55^+$), $Lin^-SCA1^+VAP1^-$ ($VAP1^-$) and $Lin^-SCA1^+VAP1^+$ ($VAP1^+$), $Lin^-SCA1^+ABCG1^-$ ($ABCG1^-$) and $Lin^-SCA1^+ABCG1^+$ ($ABCG1^+$), $Lin^-SCA1^+CD34^+CD142^-$ ($CD142^-$) and $Lin^-SCA1^+CD34^+CD142^+$ ($CD142^+$), and $Lin^-SCA1^+CD142^-ABCG1^-$ ($CD142^-ABCG1^-$) and $Lin^-SCA1^+CD142^+ABCG1^+$ ($CD142^+ABCG1^+$). The measurements were acquired using Diva software supplemented on the Becton Dickinson FACS Aria II sorter and analysed using Kaluza analysis software. The mouse sorting experiments were performed two to four times (Extended Data Fig. 3b: $CD55^+$ ASPCs = 8.7% ± 1.9%, $VAP1^+$ ASPCs = 37.0% ± 4.1%, $CD142^+$ $ABCG1^+$ ASPCs = 2.2% ± 0.7%, $CD142^+$ ASPCs = 9.6% ± 1.3%; Fig. 4e: lean subcutaneous $CD142^+ABCG1^+$ ASPCs = 2.2% ± 0.7%, obese subcutaneous $CD142^+ABCG1^+$ ASPCs = 6.6% ± 2.3%, lean visceral $CD142^+ABCG1^+$ ASPCs = 8.4% ± 2.1%, obese visceral $CD142^+ABCG1^+$ ASPCs = 13.7% ± 2.1%); all stemming from male mice.

FACS-based cell isolation of human cells. The isolated single cell suspension was diluted to 1×10^7 cells/ml with FACS buffer (DPBS −/−) with 1% human platelet serum (Cook Medical #G34936 PL-NH-500) and the following fluorophore-conjugated antibodies (Supplementary Table 17) were added: anti-human CD31-AF488, anti-human CD45-AF488 for selecting the Lin^- population, and anti-human CD142-PE (BioLegend #303110, #304017 and #365204, respectively) for separating the $CD142^+$ and $CD142^-$ populations. 7-Aminoactinomycin D (7-AAD) (Beckman Coulter #A07704) was used for assessing viability and Syto40 (Molecular Probes #S11351) was used for discerning nucleated cells. The cells were then processed as detailed above. The human ex vivo experiments were replicated for three different individuals (individuals 1 and 5, individual 4 not shown) (Extended Data Fig. 6a and Fig. 3a, b, $CD142^+$ ASPCs = 3.10% ± 0.85%). The human in vitro experiments were performed for four different individuals (individuals 1 to 4) (Extended Data Fig. 6c and Fig. 3c, d, $CD142^+$ ASPCs = 3.38% ± 0.64%). The qPCR gene expression assay was performed on ex vivo and differentiated human $CD142^+$ ASPCs and $CD142^-$ ASPCs from three different individuals (individual 1, individual 4 and individual 5) (Extended Data Fig. 6b). The 3′ RNA-seq gene expression assay (see below) was performed on in vitro and differentiated human ASPCs, $CD142^+$ and $CD142^-$ ASPCs from four distinct individuals (individuals 1 to 4).

Ex vivo adipogenic differentiation of mouse cells. The same number of cells was sorted directly into flat-bottom microscopy-adapted cell culture plates (Corning #353219), cultured to confluence in high glucose DMEM medium (Gibco #61965026) supplemented with 10% FBS and 1% penicillin–streptavidin, and treated with a single dose of white adipocyte differentiation induction cocktail (0.5 mM 3-isobutyl-1-methylxanthine (IBMX, Sigma #15879), 1 μM dexamethasone (Sigma #D2915), 170 nM insulin (Sigma #19278)), followed by maintenance treatment (1 μg/ml insulin) every 48 h. The culture was carried out until day 8, 9 or 10 after induction, at which point cells were stained for imaging or collected for RNA extraction.

Ex vivo adipogenic differentiation of human cells. The same number of cells was sorted directly into flat-bottom microscopy-adapted cell culture plates in high glucose MEMalpha medium (Gibco #32561037) supplemented with 5% human platelet serum and 50 μg/ml Primocin (InvivoGen #ant-pm-1). The cells at confluence were treated with induction cocktail (high glucose DMEM, 10% FBS, 50 μg/ml Primocin, 0.5 mM IBMX, 1 μM dexamethasone, 1.7 μM insulin, 0.2 mM indomethacin) for 7 days, followed by maintenance cocktail treatment (high glucose DMEM, 10% FBS, 50 μg/ml Primocin, 1.7 μM insulin) for another 7 days. For expanded human ASPCs, TrypLE Select reagent (Gibco #12563011) was used to collect the cells from the cell culture plates.

Quantification of in vitro differentiation (mouse and human). For experiments shown in Fig. 2c–f, h, i and Extended Data Figs 3d, e, 5d, e, h, differentiated cells were fixed with 4% formaldehyde before staining with Hoechst (nuclei) and LD540 (lipid droplets). Images were taken per well with an automatic imaging system (Operetta, Perkin Elmer) and analysed for lipid droplet content using the Harmony software. The differentiation quantification experiments were repeated at least 3 times, with multiple independent wells per experiment. At least 15 images per well (96-well plate) were acquired. One well is represented (Fig. 2c and Extended Data Fig. 3d), and quantification of 4 or 5 independent wells is shown (Fig. 2d and Extended Data Fig. 3e).

For experiments shown in Fig. 3a–d and Extended Data Figs 5a, b, 6j, k, once the cells differentiated fully, they were stained with live fluorescence dyes: Bodipy (boron-dipyrromethene, Invitrogen #D3922) for lipids and Hoechst for nuclei. Cells were incubated with the dyes in FluoroBrite phenol red-free DMEM medium (Gibco #A1896701) supplemented with 10% FBS and 1% penicillin–streptavidin for 30 min at 37 °C in the dark, washed twice with warm PBS and imaged in FluoroBrite medium using a Leica DMI4000 wide filed microscope (objectives: PL-S-APO 5×/0.15, PL-S-APO 10×/0.30, HC-PL-APO 20×/0.70) or a Zeiss

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LSM700 confocal inverted microscope (objectives: EC Plan-Neofluar 10×/0.30, Plan-Apochromat 20×/0.80). To accurately estimate and represent differences in adipose differentiation, a quantification algorithm for image treatment was developed in collaboration with the EPFL BIOP imaging facility. In brief, image analysis was performed in ImageJ/Fiji, lipid droplets (yellow) and nuclei (blue) images were filtered using a Gaussian blur (sigma equal to 2 and 3, respectively) before an automatic thresholding. The automatic thresholding algorithm selections were chosen on the basis of visual inspection of output images. At least 7 images per well (96-well plate) were acquired. Images of one technical replicate (field of view) are represented in Fig. 3a, and quantification for 2 individuals is represented in Fig. 3b for ex vivo experiments. Images of one technical replicate are represented in Fig. 3c, and quantification for 4 individuals is represented in Fig. 3d for in vitro experiments. Data points shown in Fig. 3b, d correspond to technical replicates (fields of view). The experiments were performed for at least three individuals.

Titration experiment. An equal number of Lin$^-$SCA1$^+$CD142$^-$ABCG1$^-$ and Lin$^-$SCA1$^+$CD142$^+$ABCG1$^+$ cells (see 'FACS-based cell isolation of mouse cells') were plated in 10-cm collagen I (Corning #354249)-coated cell culture plates and cultured for expansion in high glucose DMEM medium supplemented with 10% fetal bovine serum and 1% penicillin–streptavidin, refreshed every 48 h. At 90% confluence, the CD142$^-$ABCG1$^-$ and CD142$^+$ABCG1$^+$ populations were mixed at different ratios (ranging from 0 to 100%) in a 96-well plate at a density of 25,000 cells per well. Cells were differentiated as described in 'Ex vivo adipogenic differentiation of mouse cells'. The experiment was replicated twice, with $n = 4$ independent wells each. At least 15 images per well (96-well plate) were acquired. One independent well is represented in Fig. 2e, and quantification of four independent wells is represented in Fig. 2f.

siRNA-mediated knockdown experiments. For each gene, a pool of 2–4 siRNA probes (Supplementary Table 18) was reverse-transfected to total SVF (Extended Data Fig. 5c–f) or CD142$^+$ ASPCs (Aregs) (Fig. 2h, i). Seventy-five thousand cells/cm^2 were plated with 100 nM of a given siRNA dissolved in 1.5% Lipofectamine RNAiMAX (Invitrogen #13778150) in Opti-MEM I reduced serum medium (Invitrogen #31985062), and high glucose DMEM medium supplemented with 2.5% FBS (w/o penicillin–streptavidin). After 24 h, the medium was changed to high glucose DMEM medium supplemented with 10% FBS and 1% penicillin–streptavidin and after 48 h the cells were collected for determining knockdown efficiency. The experiment was replicated twice, with $n = 6$–8, 2 biological replicates and 3–4 independent wells each. At least 15 images per well (96-well plate) were acquired. One independent well is represented in Extended Data Fig. 5d, and quantification of $n = 6$–8 independent wells from two biological replicates and two independent experiments (S1 and S2) is represented in Extended Data Fig. 5e, f.

Transwell experiments. The corresponding genes were knocked down in Aregs on the transwell inserts (Corning #3381) as described in 'siRNA-mediated knockdown experiments'. 24 h after transfection, the inserts were washed with PBS and changed to high glucose DMEM medium supplemented with 10% FBS and 1% penicillin–streptavidin. Afterwards, the inserts were co-cultured with CD142$^-$ ASPCs and at 48 h, both cell populations were treated with white adipocyte differentiation cocktail (see 'Ex vivo adipogenic differentiation of mouse cells') and imaged at day 6 after induction (see 'Quantification of in vitro differentiation'). The experiment was replicated twice, with $n = 6$, 2 biological replicates and 3 independent wells each. At least 15 images per well (96-well plate) were acquired. One independent well is represented in Fig. 2h, and quantification of $n = 6$–8 independent wells from two biological replicates and two independent experiments (S1 and S2) is represented in Fig. 2i.

The experiment shown in Extended Data Fig. 5a, b was performed three times, a technical replicate (two fields of views of one well) is shown in Extended Data Fig. 5a and the quantification of two biological replicates is represented in Extended Data Fig. 5b; data points correspond to technical replicates (fields of view of one well). The experiments were imaged at day 8, 9 or 10.

In vivo differentiation. In Fig. 4f–h, 10^6 SVF cells and 10^6 CD142$^-$ABCG1$^-$ SVF cells were re-suspended in 150 µl of Matrigel (Corning #356234) and injected subcutaneously in each flank of the same 5–6-week-old mouse. This experimental set-up minimizes inter-individual variation, and enables a direct comparison of the effect of CD142$^+$ABCG1$^+$ cell removal on adipocyte differentiation capacity in vivo. In Extended Data Fig. 7i, j, 10^6 Lin$^-$SCA1$^+$ cells and 10^6 Lin$^-$SCA1$^+$CD142$^-$ cells were injected as described for Fig. 4f–h. After 3 weeks of high-fat diet, all Matrigel plugs were excised and fixed in 4% paraformaldehyde (PFA) overnight, dehydrated and embedded in paraffin. Sections of 4 µm were stained with haematoxylin and eosin. From each plug, images of at least 3 full sections were taken and adipocyte numbers, as well as the number of nuclei, were determined with the software CellProfiler[38], as previously described[25]. Five (Fig. 4f–h and Extended Data Fig. 7h) and seven (Extended Data Fig. 7i, j) biological replicates are represented. The experiments were performed once each.

Vascularisation. The Matrigel plugs were obtained as described in 'In vivo differentiation'. After excision, the Matrigel plugs were frozen in Neg-50 Frozen Section Medium (ThermoFisher #6502). Sections of 100 µm were prepared using a Reichert cryostat microtome. Primary antibody anti-isolectin GS-IB4 Alexa Fluor-488 (Invitrogen #121411) was applied in 5% goat serum and incubated 12 h at 4 °C. The sections were then washed with PBS 3 times for 10 min, at room temperature. Finally, the sections were incubated with Hoechst (5 µg/ml), washed twice for 10 min with PBS and mounted with ProLong Diamond Antifade Mountant (Thermo Fisher #P36965). The slides were then imaged with a Leica TCS SP8 microscope. One replicate is represented in Extended Data Fig. 7k. The experiment was performed once, and included 6 biological replicates (Extended Data Fig. 7l).

RNA isolation and quantitative PCR. *RNA isolation of sorted cells.* Live cells were collected in RLT+ lysis buffer (Qiagen #1053393) and flash-frozen on dry ice. Cell lysates were homogenized with QIAshredders before RNA isolation using the RNeasy Plus Micro Kit (Qiagen #74034). Reverse transcription was performed using the QuantiTect whole transcriptome kit (Qiagen #207045), following the manufacturer's recommendations for a standard-yield reaction (2 h of amplification time), or SuperScript VILO cDNA Synthesis Kit (Invitrogen #11754050). mRNA expression was normalized to *36b4* (also known as *Rplp0*) (for all mouse experiments, if not otherwise specified) or to *Hprt1* (for the human experiments and the experiments represented in Fig. 2b and Extended Data Fig. 3c (CD142$^+$ versus CD142$^-$)).

RNA isolation of differentiated cells. mRNA was isolated and transcribed into cDNA using the MultiMACS cDNA Synthesis Kit (Miltenyi #130-094-410) or collected into Tri-Reagent (Molecular Research Center #TR118). Direct-zol RNA kit (Zymo Research #R2052) was used to extract RNA, followed by reverse transcription using the SuperScript VILO cDNA Synthesis Kit (Invitrogen). Expression levels of mRNA were assessed by real-time PCR using the PowerUp SYBR Green Master Mix (Thermo Fisher Scientific #A25743). mRNA expression was normalized to *36b4* (for mouse experiments) or to *Hprt1* or *HPRT1* (for mouse or human experiments presented in Fig. 2b (third panel) and Extended Data Fig. 3c (third panel)).

Gene expression assays in Extended Data Fig. 3c show mean values over $n = 4$ or 5 biological replicates; the experiment was performed once. Figure 2b and Extended Data Fig. 3c (CD142$^+$ versus CD142$^-$) show $n = 3$ technical qPCR replicates, from 1 sorting tube. The experiment was performed twice. Extended Data Fig. 3f shows $n = 4$ or 5 independent wells, the experiment was replicated four times. Extended Data Fig. 5c, g represents 2–3 independent wells from 2 biological replicates each ($n = 4$–6 independent wells). The experiment was performed once. Extended Data Fig. 5j represents 3 independent wells from 2 biological replicates each ($n = 6$). The experiment was replicated two times. Extended Data Fig. 6b represents three biological replicates, from three individuals (individuals 1, 4 and 5).

Bulk mRNA-seq. Bulk mRNA sequencing (Figs 2, 3 and Extended Data Figs 4, 6) was performed as previously described[39,40]. In brief, 50 ng of total RNA from each sample was reverse transcribed in a 96-well plate using Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific #EP0753) with individual oligo-dT primers, featuring a 6-nt-long multiplexing barcode, and template switch oligo (Microsynth Custom made). Specifically, each oligo-dT primer is biotinylated and has the following structure: 5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT[BC6] [N15] [T30] VN-3′, in which [BC6] represents a 6-nt barcode that is specific to each well and N15 represents a stretch of random nucleotides forming a unique molecular identifier (UMI). Thus, for each well, we have generated a unique combination of barcodes and UMIs to identify each well (sample) and transcript. Next, all the samples were pooled together, purified using the DNA Clean and Concentrator kit (Zymo Research #D4014), and treated with exonuclease I (NEB or New England BioLabs #M0293S). The full-length cDNA library was amplified using a single primer and purified with Agencourt AMPure XP beads (Beckman Coulter A63881). The sequencing library was prepared by tagmentation of 10 ng full-length cDNA with a Tn5 transposase made in house, at 55 °C for 9 min[41]. Tagmented DNA was purified with the DNA Clean and Concentrator kit and PCR amplified using NEBNext High-Fidelity 2X PCR Master Mix (NEB or New England BioLabs #M0541S) with an i7 adaptor identical to Illumina Nextera and custom i5 (Microsynth custom made). The PCR reaction was then purified twice with AMPure beads or Agencourt AMPure XP beads and the average fragment size of the library was evaluated using a Fragment Analyzer (Advanced Analytical) before paired-end sequencing with NextSeq 500 (Illumina).

Analysis of bulk mRNA-seq data. For conventional bulk mRNA sequencing analysis, FASTQ files containing 100-bp-long single-end sequenced tags (reads) from four replicates of Lin$^-$CD29$^+$CD34$^+$SCA1$^+$RFP$^+$ and Lin$^-$CD29$^+$CD34$^+$SCA1$^+$RPF$^-$ each were analysed. Reads from each sample were trimmed and filtered using prinseq 0.20.3[28] with the parameters '-custom_params 'A 70%;T 70%;G 70%;C 70%' -trim_tail_left 36 -trim_tail_right 36 -lc_method dust -lc_threshold 45 -min_gc 1 -out_format 3' and cutadapt 1.5[29] with the parameters '-m 36 -q 20' and the Nextera adaptor sequence. The retained tags were evaluated using FastQC v.0.11.2 and aligned to the Ensembl 84[30] gene annotation of the NCBI38/mm10 mouse genome using STAR 2.4.0g[31] with the parameters

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

'–runThreadN 4–runMode alignReads–outFilterType BySJout–outFilterMultimapNmax 20–alignSJoverhangMin 8 –alignSJDB overhangMin 1–outFilterMismatchNmax 999–outFilterMismatchNoverLmax 0.04–alignIntronMin 20–alignIntronMax 1000000–alignMatesGapMax 1000000 g–outSAMtype BAM SortedByCoordinate'. The number of tags per gene was calculated using htseq-count 0.6.0[32] with the parameters 'htseq-count -m intersection-nonempty -s no -a 10 -t exon -i gene_id'. For each gene, expression estimates per gene were expressed as log-transformed counts per million (log_cpm), by dividing total tags per gene by the total number of gene-aligned reads per cell and taking the $\log(x+1)$ of this value.

To analyse the data from barcoded mRNA sequencing (Areg-related experiments in mouse and human, Figs 2, 3 and Extended Data Figs 4, 6)[40], the following workflow was used. Reads from barcoded mRNA-seq experiments have two barcodes, corresponding to the two levels of multiplexing. The first one is common to standard protocols and is used to separate the libraries. The second is specific to the barcoded mRNA-seq protocol and is used to separate the multiplexed samples from the bulk data. The first demultiplexing step was performed with the Illumina BaseSpace platform, and the second was performed using custom scripts. FASTQ files containing 62-bp-long single-end sequenced tags (reads) from at least four biological replicates (4 biological replicates, each with 1–3 independent wells) per ASPC population and sample were analysed: 6 ASPC replicates D0, 4 ASPC samples D12, 8 CD142$^-$ ASPC replicates D0, 7 CD142$^-$ ASPC 5 h, 7 CD142$^-$ ASPC 24 h, 4 CD142$^-$ ASPC D12, and 4 replicates each for CD142$^+$ ASPC D0, 5 h, 24 h and D12 (Fig. 2g and Extended Data Fig. 4). Reads from each sample were trimmed and filtered using prinseq 0.20.3[28] with the parameters '-custom_params 'A 70%;T 70%;G 70%;C 70%' -trim_tail_left 36 -trim_tail_right 36 -lc_method dust -lc_threshold 45 -min_gc 1 -out_format 3' and cutadapt 1.5[29] with the parameters '-m 36 -q 20' and the Nextera adaptor sequence. The retained tags were evaluated using FastQC v.0.11.2 and aligned to the Ensembl 84[30] gene annotation of the NCBI38/mm10 mouse genome using STAR 2.4.0g[31] with the parameters '–runThreadN 4–runMode alignReads–outFilterType BySJout–outFilterMultimapNmax 20–alignSJoverhangMin 8 –alignSJDB overhangMin 1–outFilterMismatchNmax 999–outFilterMismatchNoverLmax 0.04–alignIntronMin 20–alignIntronMax 1000000–alignMatesGapMax 1000000 g–outSAMtype BAM SortedByCoordinate'. The number of tags per gene was calculated using htseq-count 0.6.0[32] with the parameters 'htseq-count -m intersection-nonempty -s no -a 10 -t exon -i gene_id'.

Genes with a count per million greater than 2 in at least 4 samples were retained, providing a filtered dataset of 26,159 expressed genes across 52 samples. The samples were of high quality, as assessed by the number of aligned reads and detected genes per cell and by correlations with the P3 single-cell data (Extended Data Fig. 4c, d). Raw counts were normalized using mean-variance modelling at the observational level, as implemented in the voom() function in limma_3.30.4[42] and further using combat() in sva_3.22.0[43] to adjust for batch effects. Differential expression was computed on the normalized values using the limma_3.30.4 pipeline at an FDR of 0.05 and fold-change cutoff of 2. We detected a large number of changes that were induced upon adipogenic differentiation in all three cell fractions (CD142$^+$, CD142$^-$ and ASPCs) (Extended Data Fig. 4e, Supplementary Table 11), each of which had over 3,500 significantly differentially expressed genes (FDR 0.05, fold-change 2). Across time points, the highest number of differences was observed between Aregs and CD142$^-$ cells or ASPCs (Extended Data Fig. 4e). The high transcriptional similarity between CD142$^-$ cells and ASPCs (Extended Data Fig. 4e, f) was expected based on the low fraction of Aregs (generally $<10\%$) contained in ASPCs. Most differences arose early upon plating (after 5 h and 24 h), but dampened after adipogenic differentiation (Extended Data Fig. 4e). Heat maps displaying row-normalized (z-score transformation) expression (log_cpm) values were generated using pheatmap_1.0.8 and the parameters 'clustering_distance_rows = "correlation",cluster_cols = F, clustering_method = "average", scale = "row"'. Extended Data Fig. 4g contains all genes that were expressed at significantly lower levels after adipogenic differentiation (day 12, D12) in CD142$^+$ ASPCs compared to both CD142$^-$ ASPCs and all ASPCs; Fig. 2g contains all genes that were expressed at significantly higher levels in CD142$^+$ ASPCs compared to CD142$^-$ ASPCs after sorting (day 0, D0) and after culturing (5 h and day 1, D1); Extended Data Fig. 4k contains all genes that were expressed at significantly higher levels in CD142$^+$ ASPCs compared to CD142$^-$ ASPCs after sorting (day 0, D0); Extended Data Fig. 4n shows the expression of selected endothelial marker genes. All correlations were calculated based on log_cpm values, with the function cor() and the parameters 'method="Spearman"'.

The analysis of the human barcoded mRNA-seq data (displayed in Fig. 3 and Extended Data Fig. 6) was performed in a manner analogous to that described above for mouse data, but reads were aligned to the Ensembl 84[30] gene annotation of the GRCh38/hg20 human genome instead. Figure 3e contains only the top 30 genes that were expressed at significantly higher levels (fold-change = 2, FDR = 0.1) in CD142$^-$ ASPCs compared to CD142$^+$ ASPCs, and Extended Data Fig. 6g contains all genes that were expressed at significantly higher levels (fold-change = 2, FDR = 0.1) in CD142$^+$ ASPCs compared to CD142$^-$ ASPCs.

Immunofluorescence. Mice were anaesthetized with isoflurane and perfused with PBS (5 min) followed by 4% PFA (paraformaldehyde, electron microscopy grade (VWR #100504-858) (5 min). The subcutaneous fat pads were dissected and post-fixed in 4% PFA for 2 h at 4 °C upon gentle shaking. Next, the tissue was washed with PBS and incubated with 30% sucrose for 24 h at 4 °C upon gentle shaking. Cryoblocks were prepared using a Cryomatrix (Thermo Fisher Scientific #6769006) and 30-μm sections were generated using a Leica CM3050S cryostat at −30 °C. The sections were deposited onto glass slides and incubated in −20 °C acetone for 10 min, after which they were dried at room temperature for 10 min, rehydrated in PBS and blocked with 10% goat serum (Invitrogen #31872) supplemented with 0.3% TritonX100 (Sigma #T9284) for 1 h. Primary antibodies (Supplementary Table 17, hamster anti-mouse CD31, BioRad #MCA1370A, rabbit anti-mouse CD142, SinoBiological #50413-R001, rat anti-mouse SCA1, BioLegend #122501) were applied in 1% goat serum with 0.3% TritonX100 and incubated for 24 h at 4 °C. The sections were then washed with PBS three times for 10 min at room temperature. Secondary antibodies (Supplementary Table 17, goat anti-hamster AF-546, Molecular Probes #A21111, donkey anti-rabbit AF-488, Molecular Probes #A21247, goat anti-rat AF-647, Molecular Probes #A21206) were applied in PBS with 0.3% TritonX100 and incubated in the dark for 1 h at room temperature, followed by three 10-min washes with PBS at room temperature in the dark. Finally, the sections were incubated with Hoechst (5 μg/ml), washed two times for 10 min with PBS and mounted with Fluoromount G (Southern Biotech #0100-01). The slides were then imaged with a Zeiss confocal LSM700 microscope (objectives: EC Plan-Neofluar 10×/0.30, Plan-Apochromat 20×/0.80, Plan-Apochromat 40×/1.30). The results presented in Fig. 4a–d were replicated in at least three independent experiments. We note that we also verified that the signal we detected is not the result of autofluorescence of the adipose tissue (Extended Data Fig. 7a) or from unspecific binding of secondary antibodies (Extended Data Fig. 7b). We also showed that the perivascular staining is not an artefact caused by perfusing the animals with 4% PFA (Extended Data Fig. 7c).

Annotation and functional enrichment analyses. Gene annotations were obtained from Ensembl 84 (mar2016.archive.ensembl.org)[30], either by direct download or through biomaRt_2.30.0[44]. Enrichment analysis was performed using the webserver Enrichr with default parameters[45] and annotations provided by KEGG[46], Wikipathways[47], or gene expression data provided by the GTEx project[48]. For Extended Data Fig. 1g, the top 10 enriched terms were considered. Only highlights are displayed in the figures, and full results are available in Supplementary Tables 3, 8, 12 and 16.

Other computational analyses and data processing remarks. All computational analyses were performed using R version 3.3.2 and Bioconductor version 3.4. All t-tests and Wilcoxon rank-sum tests were unpaired and two-sided, if not otherwise specified. All box plots were generated and displayed in R, using the boxplot() function with default parameters. The median value is indicated with a black line, and a coloured box (hinges) is drawn between the 1st and 3rd quartiles (interquartile range, IQR). The whiskers correspond to approximately 1.5× interquartile range (±1.58 interquartile range divided by the square root of n) and outliers are drawn as individual points. All bean plots are generated and displayed in R, using the beanplot() (package beanplot) function with default parameters ('bw = "SJ-dpi", kernel = "gaussian", cut = 3, cutmin = -Inf, cutmax = Inf'). The mean value is indicated with the widest black line, and individual values are indicated with narrower lines. Kernel density estimates are displayed in colours. All bar plots display mean values as centres and the standard deviation as error bars. All included microscopy images and macroscopic images are representative. Exact P values are provided in the Supplementary Table 19; only the P value ranges are provided in the figure legends.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. Sample scripts used to process the data are available at https://github.com/DeplanckeLab/Areg.

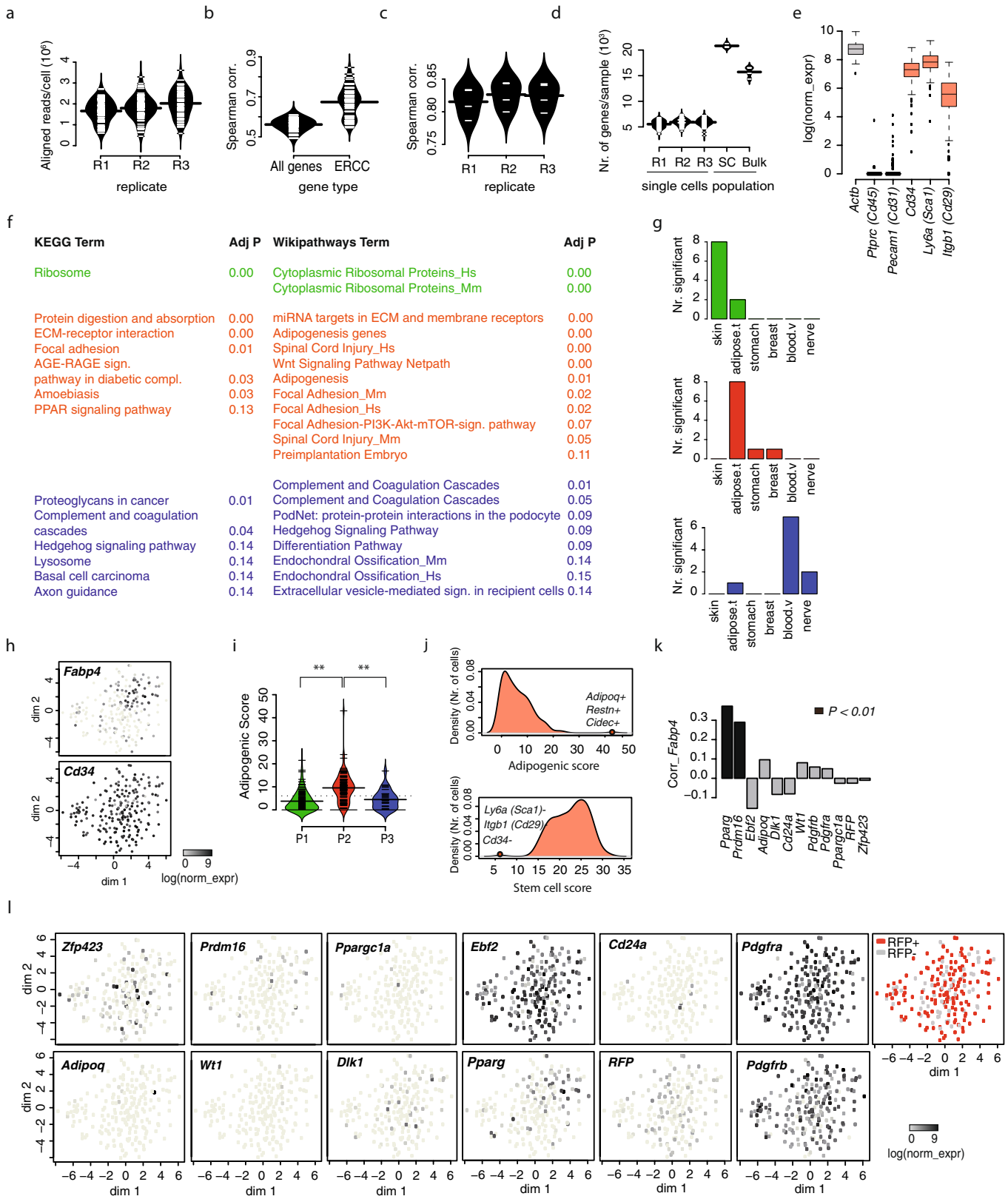
Data availability. All raw and processed RNA-seq data have been uploaded in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) with the accession numbers E-MTAB-5785, E-MTAB-5818, E-MTAB-5802, E-MTAB-5787 and E-MTAB-6677. Microscopy images are available from the corresponding authors upon reasonable request.

26. Johansson, T. et al. Building a zoo of mice for genetic analyses: A comprehensive protocol for the rapid generation of BAC transgenic mice. *genesis* **48**, 264–280 (2010).
27. Rosenwald, M., Perdikari, A., Rülicke, T. & Wolfrum, C. Bi-directional interconversion of brite and white adipocytes. *Nat. Cell Biol.* **15**, 659–667 (2013).
28. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

29. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
30. Yates, A. et al. Ensembl 2016. *Nucleic Acids Res.* **44**, D710–D716 (2016).
31. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
32. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
33. Andrews, T. S. & Hemberg, M. Modelling dropouts allows for unbiased identification of marker genes in scRNASeq experiments. Preprint at https://www.biorxiv.org/content/early/2016/07/21/065094 (2016).
34. Van Der Maaten, L. J. P. & Hinton, G. E. Visualizing data using *t*-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
35. Kiselev, V. Y. et al. SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14**, 483–486 (2017).
36. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res.* **5**, 2122 (2016).
37. Scialdone, A. et al. Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods* **85**, 54–61 (2015).
38. Carpenter, A. E. et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
39. Pradhan, R. N. et al. Dissecting the brown adipogenic regulatory network using integrative genomics. *Sci. Rep.* **7**, 42130 (2017).
40. Alpern, D., Gardeux, V., Russeil, J. & Deplancke, B. Time- and cost-efficient high-throughput transcriptomics enabled by Bulk RNA Barcoding and sequencing. Preprint at https://www.biorxiv.org/content/early/2018/01/30/256594 (2018).
41. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protocols* **9**, 171–181 (2014).
42. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
43. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
44. Drost, H.-G. & Paszkowski, J. Biomartr: genomic data retrieval with R. *Bioinformatics* **33**, 1216–1217 (2017).
45. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
46. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
47. Kelder, T. et al. WikiPathways: building research communities on biological pathways. *Nucleic Acids Res.* **40**, D1301–D1307 (2012).
48. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

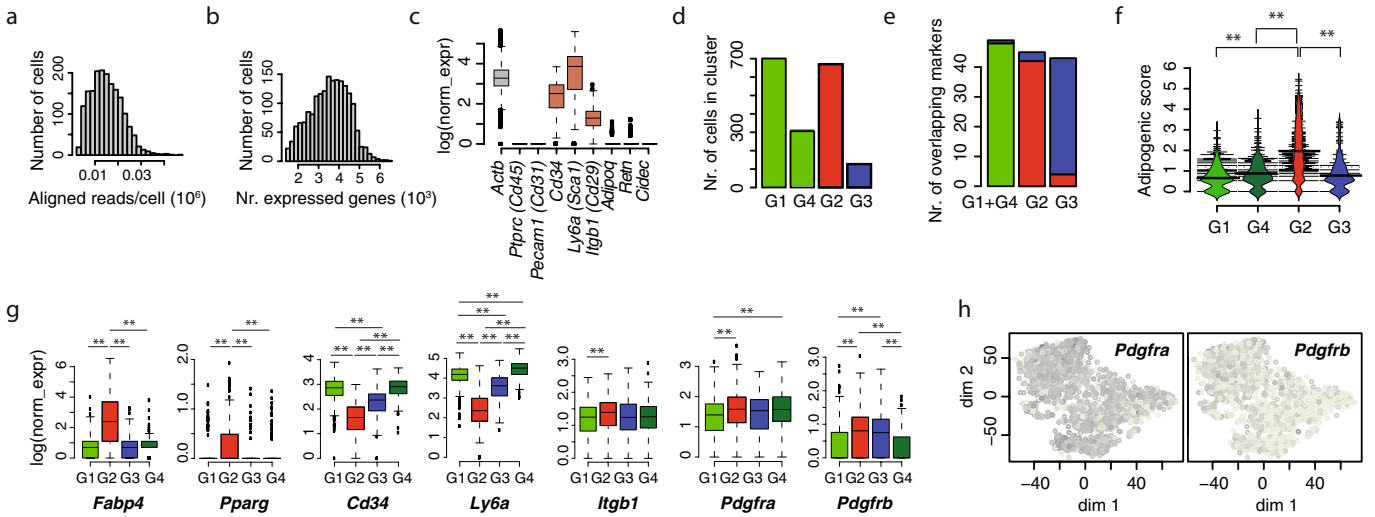


Extended Data Fig. 1 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

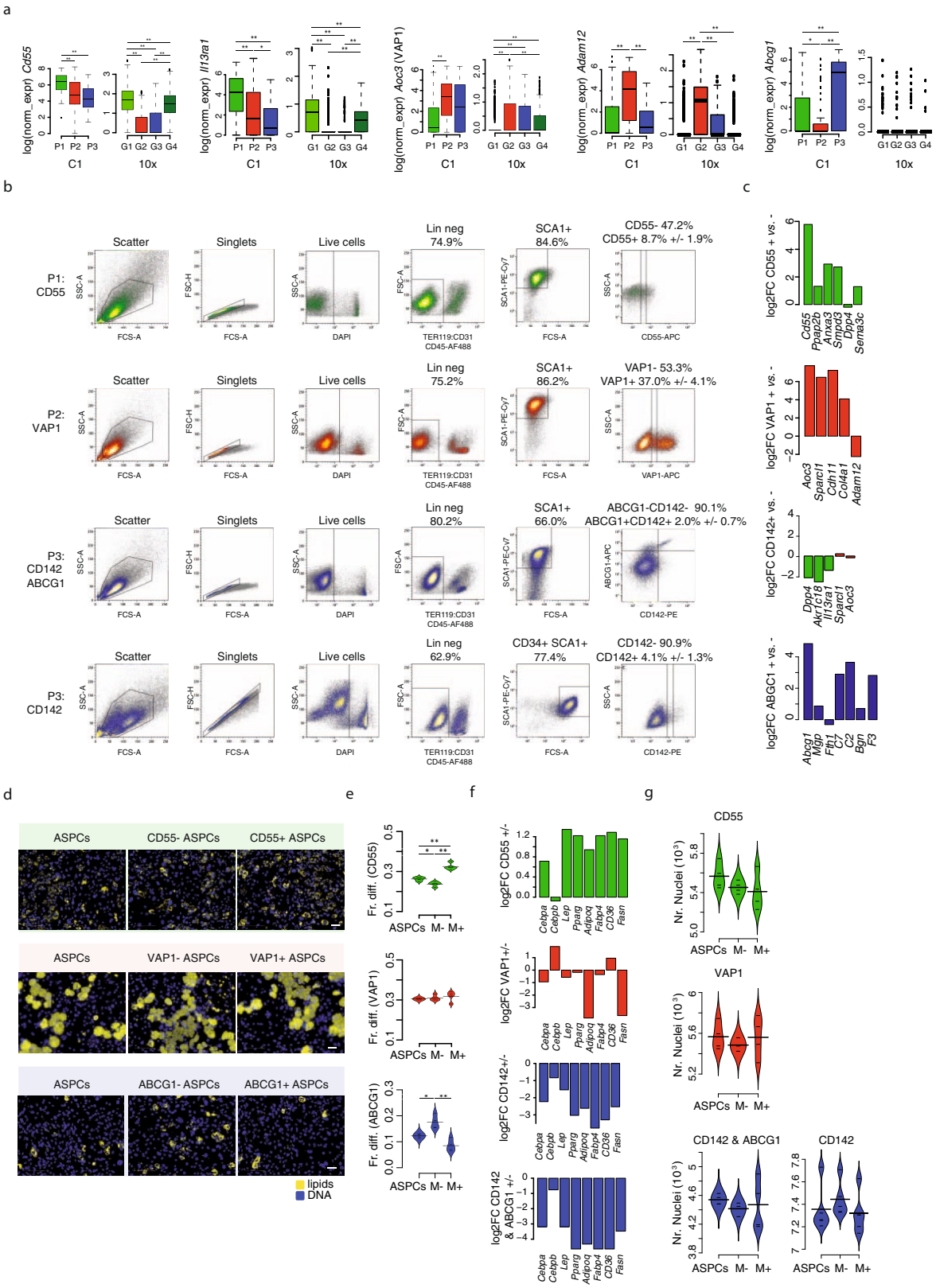
Extended Data Fig. 1 | ScRNA-seq reveals the heterogeneity of ASPCs (Fluidigm C1). a, Number of aligned reads per cell for each of the three Fluidigm C1 (C1) scRNA-seq experiments. R1, $n = 74$; R2, $n = 71$; R3, $n = 63$ single cells from three independent biological replicates, each derived from a pool of mice. **b**, Correlations (corr, Spearman's rho) between expression in single cells across all genes or only the artificial RNA spike-ins (ERCC). **c**, Correlation (Spearman's rho) between expression in merged single cell (per individual biological replicate, R1–R3) and bulk population ($n = 4$ biological replicates) samples. **d**, Number of expressed genes/sample in each of the three biological scRNA-seq replicates (R1–R3), as well as the merged individual replicates (SC, $n = 3$) or the bulk population control samples (bulk, $n = 4$). **e**, Expression of ubiquitous *Actb*, negative (*Ptprc* and *Pecam1*) and positive (*Cd34*, *Ly6a* and *Itgb1*) FACS markers. **f**, **g**, Pathways[47] significantly enriched (**f**) and top 10 significantly enriched GTEx[48] tissue samples (**g**) among the top 200 genes specific to one of the three populations (P1, green; P2, red; P3, blue). Full enrichment results are listed in Supplementary Table 3. **h**, *t*-SNE 2D maps of all analysed cells (C1), highlighting the expression of the stem cell marker *Cd34* and the adipogenic marker *Fabp4* (black). **i**, Bean plots showing the distribution of adipogenic scores across cells belonging to one of the three populations (P1, $n = 83$; P2, $n = 96$; P3, $n = 29$ cells). $^{**}P \leq 0.01$, Wilcoxon rank-sum test. **j**, Stem cell (lower panel) and adipogenic (upper panel) score distributions across all analysed cells, highlighting a single outlier cell not expressing—or expressing very low levels of—*Itgb1* (CD29), *Cd34* and *Ly6a* (SCA1) as well as very high levels of the mature adipogenic markers *Adipoq*, *Restn* and *Cidec*. **k**, Correlation (Spearman's rho) between the expression of 12 (pre-)adipogenesis-related genes and the expression of *Fabp4*; significant (Bonferroni multiple testing adjusted $P \leq 0.01$) correlations in black. **l**, *t*-SNE 2D maps of all analysed cells, highlighting the expression of *RFP* and various genes previously used to mark adipogenic precursors or pre-adipocytes (black); microscopic RFP status also displayed. In **e**, **h**, **j**, **l**, $n = 208$ cells from three independent biological experiments, each based on a pool of mice.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 2 | ScRNA-seq reveals the heterogeneity of ASPCs (10x Genomics). **a**, **b**, Number of aligned reads per cell (**a**) and number of expressed genes per cell (**b**) for the scRNA-seq experiment performed using the 10x Genomics Chromium instrument. **c**, Expression of ubiquitous *Actb*, negative (*Ptprc* and *Pecam1*), positive (*Cd34*, *Ly6a* and *Itgb1*) FACS markers and adipocyte markers (*Adipoq*, *Retn* and *Cidec*). **d**, Number of cells in each of the four 10x Genomics populations (G1–G4). **e**, Percentage of the top 100 marker genes of the 10x Genomics populations (G1 and G4, G2, and G3) that overlap the top 100 marker genes of the C1 cell populations (P1–P3). **f**, Bean plots showing the distribution of adipogenic scores across cells that belong to one of the four 10x Genomics populations (G1–G4). Only *P* values for G2 comparisons are marked. **g**, Box plots showing the distribution of log(normalized expression) values of pre-adipogenic and ASPC markers across the four 10x Genomics populations (G1–G4). $*P \leq 0.05$, $**P \leq 0.01$, Wilcoxon rank-sum test. In **e**–**g**, G1, $n = 699$; G2, $n = 664$; G3, $n = 122$; and G4, $n = 319$ cells. **h**, *t*-SNE 2D maps of all analysed cells (10x Genomics), highlighting the expression of the pre-adipogenic markers *Pdgfra* and *Pdgfrb* (black). In **c**, **h**, $n = 1{,}804$ cells from one biological experiment based on a pool of mice.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

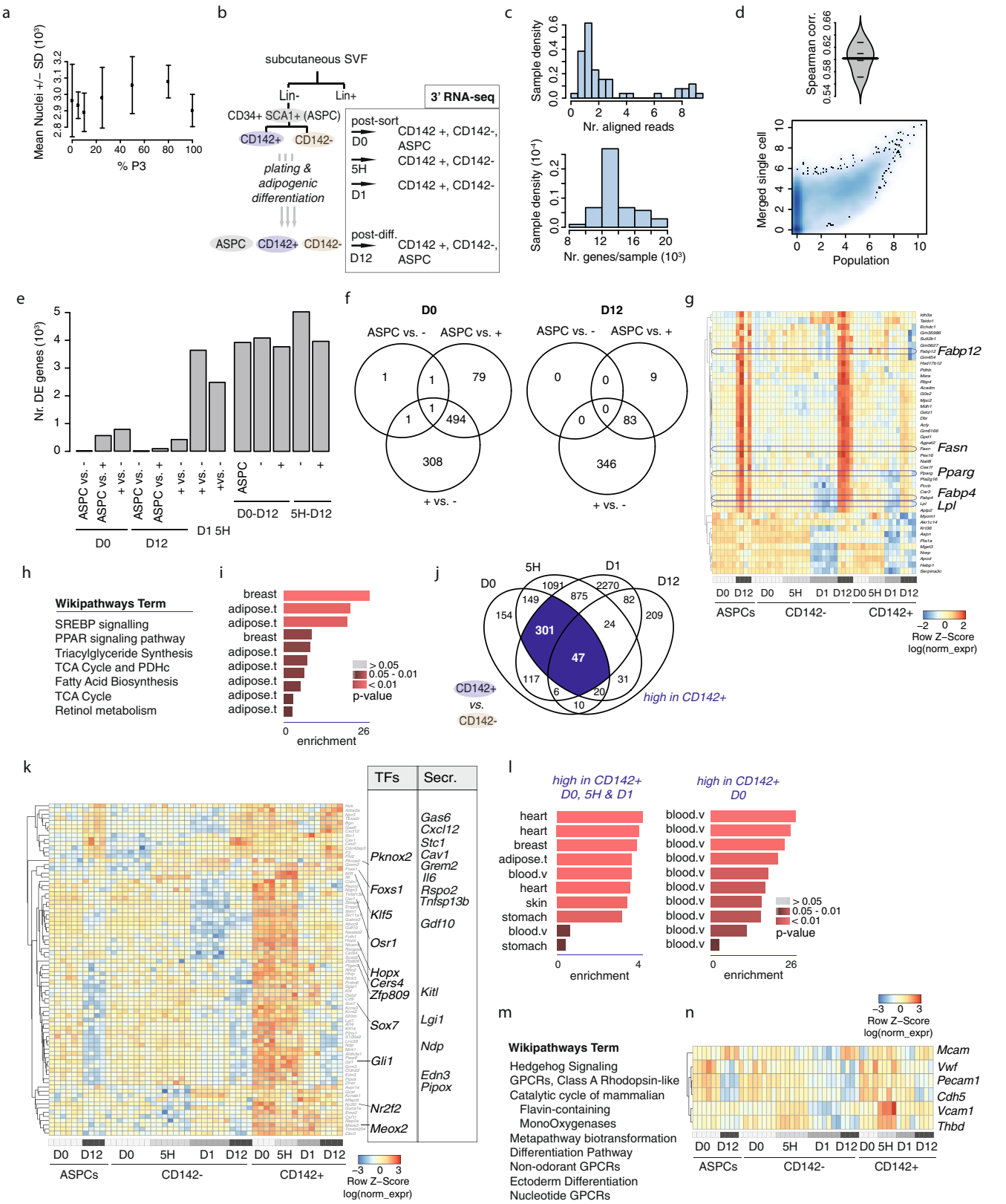


Extended Data Fig. 3 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 3 | Three ASPC subpopulations with distinct adipogenic differentiation capacity. a, Box plots showing the distribution of log(normalized expression) values for surface markers with available FACS-grade antibodies selected for subpopulation follow-up: CD55 and IL13RA1 (P1, G1 and G4), VAP1 and ADAM12 (P2 and G2) and ABCG1 (P3 and G3). $*P \leq 0.05$, $**P \leq 0.01$, Wilcoxon rank-sum test. P1, $n = 83$; P2, $n = 96$; P3, $n = 29$; G1, $n = 699$; G2, $n = 664$; G3, $n = 122$; and G4, $n = 319$ cells. **b**, FACS-based sorting strategy to isolate the three identified populations. **c**, qPCR-based expression fold-changes ($\log_2$(fold-change (FC) marker-positive versus marker-negative ASPCs)) for a panel of P1-, P2- and P3-specific genes. **d**, Microscopy images of distinct ASPC fractions after adipogenic differentiation; $CD55^+$ (P1 or G1 and G4), $VAP1^+$ (P2 or G2), and $ABCG1^+$ (P3 or G3). Nuclei are stained with DAPI (blue) and lipids with LD540 (yellow). Scale bar, 50 μm. **e**, Bean plots showing the distribution of the fraction of differentiated cells (Fr. diff.) per each ASPC subpopulation shown in **d**, $n = 4$ independent wells. M−, marker negative; M+, marker positive. $*P \leq 0.05$, $**P \leq 0.01$, *t*-test. **f**, qPCR-based expression fold-changes ($\log_2$(fold-change, marker-positive versus marker-negative ASPCs)) upon adipogenic differentiation for a panel of adipogenic marker genes. **g**, Bean plots showing the distribution of the mean nuclei number for the four differentiated ASPC fractions shown in **d**, **e** and Fig. 2c, d. $n = 4$ or 5 independent wells. In **d**, **e**, **g**, the experiments were repeated independently three times, yielding similar results; representative images are shown. All panels, population 1 (P1, green), population 2 (P2, red), population 3 (P3, blue).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

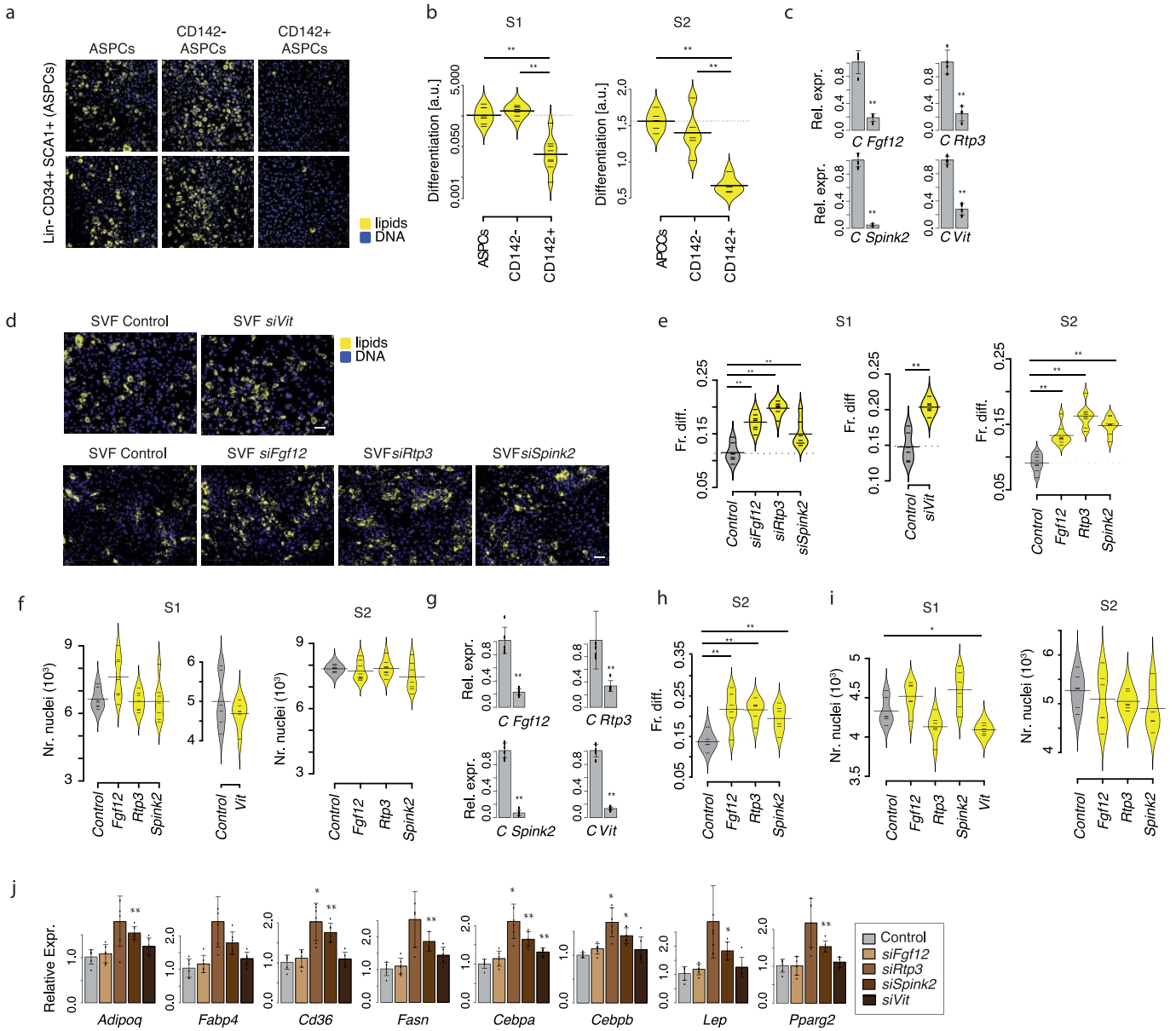


Extended Data Fig. 4 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

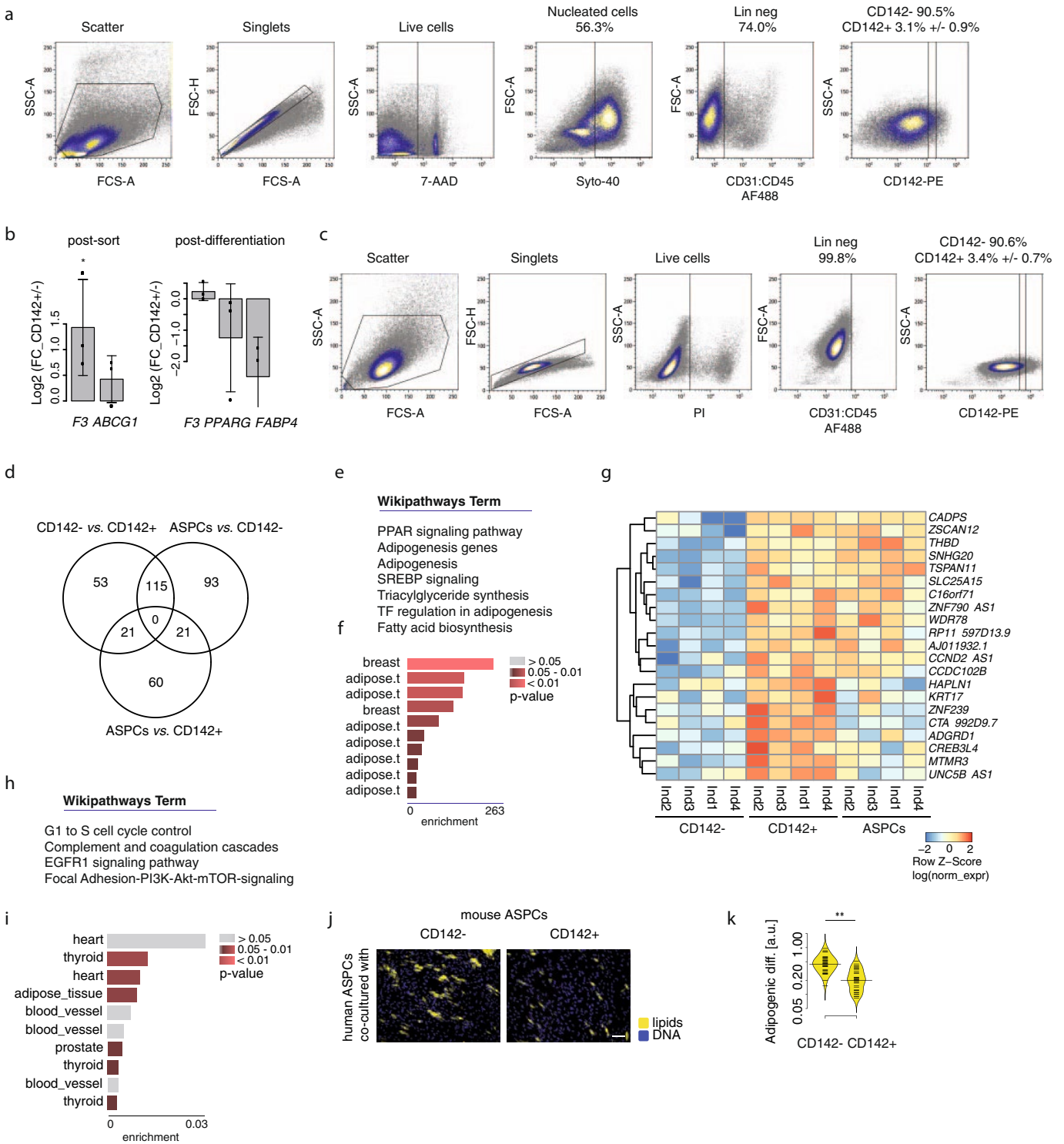
Extended Data Fig. 4 | Aregs (CD142$^+$ABCG1$^+$ ASPCs) show adipogenic inhibitory capacity. a, Mean and s.d. of number of nuclei per experiment; *x* axis represents the percentage of Lin$^-$SCA1$^+$CD142$^+$ABCG1$^+$ cells (P3, Aregs) mixed in with Lin$^-$SCA1$^+$CD142$^-$ABCG1$^-$ cells; $n = 4$ independent wells. **b**, Schematic of sample collection aiming at a detailed characterization of CD142$^+$ ASPCs versus CD142$^-$ ASPCs and all ASPCs, after sorting (D0), upon plating (5 h), after culturing (D1) and after adipogenic differentiation (D12). **c**, Number of aligned reads (top) and number of expressed genes per sample (bottom) across barcoded RNA-seq samples. **d**, Correlation (Spearman's rho) between the barcoded mRNA-seq samples after sorting, and the merged scRNA-seq P3 cells; $n = 4$ biological replicates (top). Example of the correlation (Spearman's rho) between the log(normalized expression) estimates across merged scRNA-seq P3 cells and one barcoded mRNA-seq replicate (bottom). **e**, Number of genes ($\times 10^3$) that were significantly differentially expressed (FDR 0.05, fold-change 2) in all comparisons between all ASPCs, CD142$^+$ ASPCs and CD142$^-$ ASPCs, after sorting (D0), upon plating (5 h), after culture (D1) and after adipogenic differentiation (D12) (left) as well as upon adipogenic induction (D0 versus D12, D0–D12, and 5 h versus D12, 5 h–D12), in all ASPCs, CD142$^+$ ASPCs and CD142$^-$ ASPCs (right). **f**, Venn diagram showing overlaps between significantly differentially expressed genes (FDR 0.05, fold-change 2) in CD142$^+$ ASPCs (+) versus CD142$^-$ ASPCs (−) and all ASPCs (ASPC), after sorting (D0) and after adipogenic differentiation (D12). **g**, Heat map of the expression (row-wise *z*-scores of log(normalized expression), blue-to-red) of genes significantly differentially expressed (FDR 0.05, fold-change 2) in CD142$^+$ ASPCs versus CD142$^-$ ASPCs and versus all ASPCs upon adipogenic differentiation (D12). Adipogenic marker genes are highlighted. **h**, **i**, Pathways[47] (**h**) and top 10 significantly enriched GTEx tissue samples (**i**) that are significantly enriched among the genes displayed in **f**. **j**, Venn diagram showing overlaps between significantly differentially expressed genes (FDR 0.05, fold-change 2) in CD142$^+$ ASPCs versus CD142$^-$ ASPCs at all four time points that were assessed. **k**, Heat map of the expression (row-wise *z*-scores of log(normalized expression), blue-to-red) of genes that were expressed at significantly higher levels (FDR 0.05, fold-change 2) in CD142$^+$ ASPCs versus CD142$^-$ ASPCs after sorting (D0). Genes encoding transcription factors (TFs) and secreted factors (Secr.) are highlighted. **l**, Top 10 significantly enriched GTEx tissue samples among the genes that were expressed at significantly higher levels (FDR 0.05, fold-change 2) in CD142$^+$ ASPCs versus CD142$^-$ ASPCs after sorting, plating and culture (D0, 5 h and D1, respectively) (left) as well as after sorting only (D0) (right). **m**, Pathways[47] significantly enriched among the genes displayed in **k**. For **h**, **i**, **l**, **m**, full enrichment results are listed in Supplementary Table 12. **n**, Heat map of the expression (row-wise *z*-scores of log(normalized expression), blue-to-red) of a panel of endothelial marker genes. In **g**, **k**, **n**, $n = 4$–8, 4 biological replicates, 1–3 independent wells each.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 5 | The adipogenic inhibitory capacity of Aregs is paracrine. a, Microscopy images (after adipogenic differentiation) of mouse ASPCs co-cultured in transwells with all ASPCs, CD142$^-$ ASPCs or CD142$^+$ ASPCs. Nuclei are stained with Hoechst (blue) and lipids with Bodipy (yellow). Scale bar, 50 μm. The experiments were repeated independently two times, yielding similar results; representative images are shown. **b**, Bean plots showing the extent of differentiation (arbitrary units, a.u.) per each ASPC fraction shown in **a**, as well as an additional independent biological replicate. $^{**}P \leq 0.01$, Wilcoxon rank-sum test. Experiment S1: ASPCs and CD142$^-$ ASPCs, $n = 8$; CD142$^+$ ASPCs, $n = 7$ fields of view; experiment S2: ASPCs and CD142$^-$ ASPCs, $n = 5$; CD142$^+$ ASPCs, $n = 7$ fields of view. **c**, qPCR-measured gene expression (Rel. expr.) of knockdowns of *Fgf12*, *Rtp3*, *Spink2* ($n = 4$) and *Vit* ($n = 6$) in total SVF. Two biological replicates, 2 or 3 independent wells each. **d**, Microscopy images of plated SVF cells with knockdowns of *Fgf12*, *Rtp3*, *Spink2* and *Vit*, as well as control (scramble siRNA) knockdowns, after adipogenesis. Nuclei are stained with DAPI (blue) and lipids are stained with LD540 (yellow). Scale bars, 50 μm. **e**, Bean plots showing the distribution of the fraction of differentiated SVF cells with knockdowns of *Fgf12*, *Rtp3*, *Spink2* and *Vit*, as well as control knockdowns. **f**, Bean plots showing the distribution of the mean number of nuclei for the differentiated fractions shown in **e**. In **e**, **f**, for *Fgf12*, *Rtp3* and *Spink2* in experiment S1: $n = 7$, 2 biological replicates, 3 or 4 independent wells each; in experiment S2: $n = 8$, 2 biological replicates, 4 independent wells each; for *Vit*: $n = 6$, 2 biological replicates, 3 independent wells each. **g**, qPCR-measured gene expression of *Fgf12*, *Rtp3*, *Spink2* and *Vit* in Aregs of the experiments shown in Fig. 2h, i; $n = 6$, 2 biological replicates, 3 independent wells each. **h**, Bean plots showing the distribution of the fraction of differentiated cells, as measured in the CD142- ASPCs underneath Aregs with the knockdowns of *Fgf12*, *Rtp3*, *Spink2*, and control; $n = 6$, 2 biological replicates, 3 independent wells each (independent replication of the experiment shown in Fig. 2h, i). **i**, Bean plots showing the distribution of the mean number of nuclei for the differentiated fractions shown in Fig. 2h, i (experiment S1) and Extended Data Fig. 5h (experiment S2). $n = 6$, 2 biological replicates, 3 independent wells. **j**, Expression of adipogenic markers measured using qPCR in the CD142$^-$ ASPCs underneath Aregs with knockdown of genes shown in Fig. 2h, i; $n = 6$: 2 biological replicates, 3 independent wells each; Experiments were repeated two times, yielding similar results. In **c**–**j**, $^*P \leq 0.05$, $^{**}P \leq 0.01$, *t*-test.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

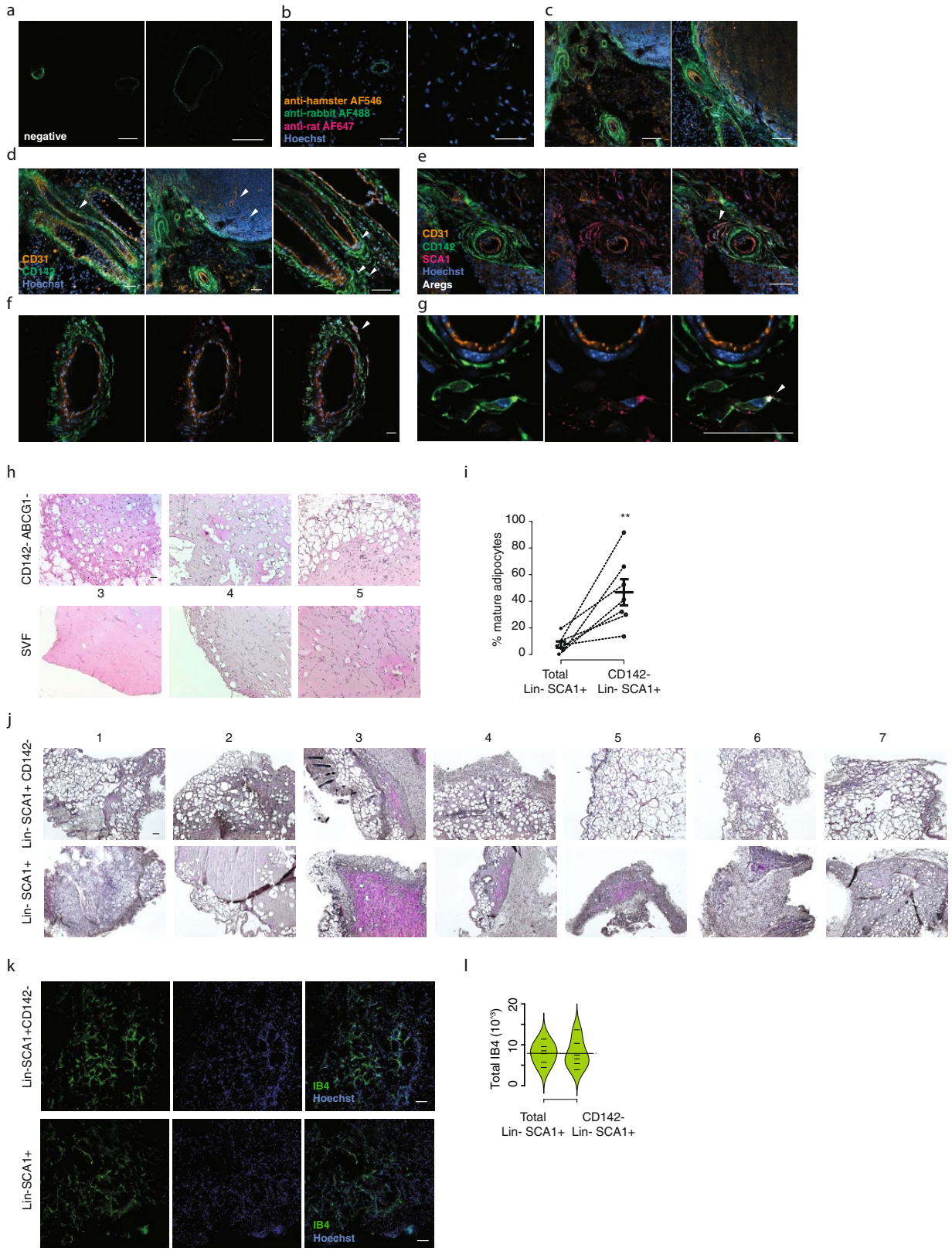


Extended Data Fig. 6 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 6 | Aregs and their inhibitory capacity are conserved in humans. a, FACS-based strategy to isolate human (one individual is shown) ex vivo $CD142^+$ ASPCs and $CD142^-$ ASPCs. **b**, qPCR-based expression fold-changes (log_2(fold change)) of $CD142^+$ ASPCs versus $CD142^-$ ASPCs for *F3* and *ABCG1* after sorting (left) and *F3*, *PPARG* and *FABP4* after differentiation (right). $n = 3$ biological replicates (distinct individuals). $^*P \leq 0.05$, one-sided paired *t*-test. **c**, FACS-based strategy to isolate the human (one individual is shown) in vitro $CD142^+$ ASPCs and $CD142^-$ ASPCs. **d**, Venn diagram showing overlaps between significantly differentially expressed genes (FDR 0.1, fold-change 2) in $CD142^+$ ASPCs versus $CD142^-$ ASPCs and all ASPCs after adipogenic differentiation (D12). **e**, **f**, Pathways[47] (**e**) and top 10 GTEx tissue samples (**f**) significantly enriched among the genes that were expressed at significantly higher levels (FDR 0.1, fold-change 2) in $CD142^-$ ASPCs versus $CD142^+$ ASPCs after adipogenic differentiation (Fig. 3e). **g**, Heat map of the expression (row-wise *z*-scores of log(normalized expression), blue-to-red) of genes that were expressed at significantly higher levels (FDR 0.1, fold-change 2) in $CD142^+$ ASPCs versus $CD142^-$ ASPCs after adipogenic differentiation (D12). **h**, **i**, Pathways[47] (**h**) and top 10 GTEx tissue samples (**i**) significantly enriched among the genes that were expressed at significantly higher levels (FDR 0.1, fold change 2) in $CD142^+$ ASPCs versus $CD142^-$ ASPCs after adipogenic differentiation (D12) and displayed in **g**. For **e**, **f**, **h**, **i**, full enrichment results are provided in Supplementary Table 16. In **d**–**i**, $n = 4$ biological replicates (distinct individuals). **j**, Microscopy images (after adipogenic differentiation) of human ex vivo ASPCs co-cultured with mouse Areg- ($CD142^+$) or Areg-depleted ($CD142^-$) ASPCs. Nuclei are stained with Hoechst (blue) and lipids with Bodipy (yellow). Scale bar, 50 μm. The experiment was performed once. **k**, Bean plots showing the extent of adipogenic differentiation (arbitrary units, a.u.) per each ASPC fraction shown in **j**. $n = 35$ fields of view. All panels except b: $^*P \leq 0.05$, $^{**}P \leq 0.01$, Wilcoxon rank-sum test.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

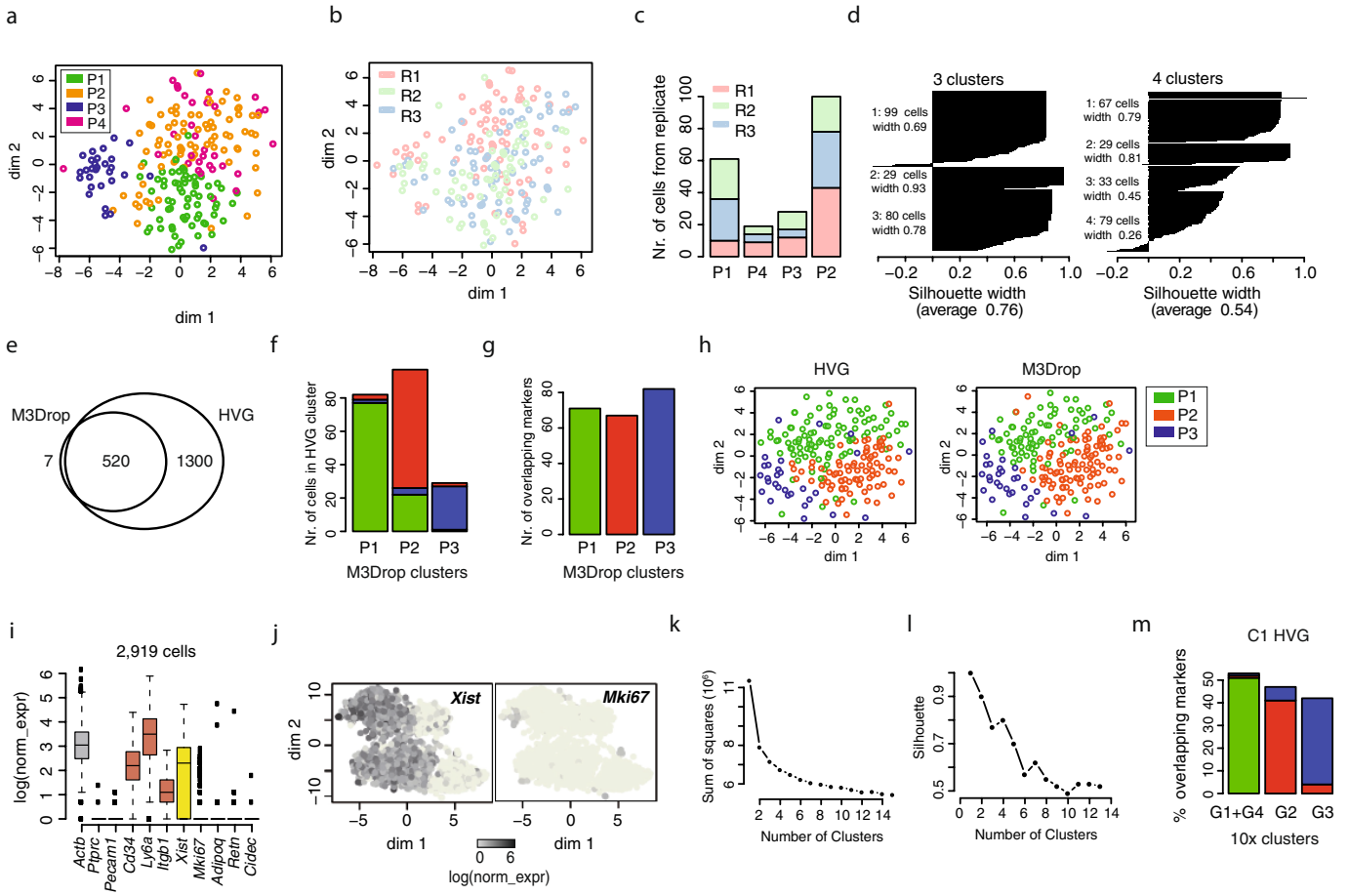


Extended Data Fig. 7 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 7 | Aregs locate proximal to the vasculature and inhibit adipogenesis in vivo. a–g, Microscopy images of the in vivo localization of CD31 (orange), CD142 (green) and SCA1 (pink) markers in mouse subcutaneous white adipose tissue. Nuclei are stained with Hoechst (blue). Scale bar, 50 μm. The experiments were repeated independently at least three times, yielding similar results; representative images are shown. Negative control (**a**) or tissue cryosections stained with the indicated secondary antibodies only (**b**). In **c**, perfused (left) versus non-perfused (right) tissue stained for the indicated markers. In **d**, arrows indicate $CD142^+$ around big vessels within the adipose parenchyma (left), outside the lymph nodes (middle) and within individual cells (right); in **e–g**, arrows indicate individual cells presenting both CD142 and SCA1 staining (white). **h**, Total SVF and Areg-depleted SVF cells ($CD142^-ABCG1^-$) were implanted into the subcutaneous adipose depot in Matrigel. Microscopy images of Matrigel plugs with total SVF or $CD142^-ABCG1^-$ SVF cells. Scale bar, 100 μm. Individuals 3 to 5 are shown, and individuals 1 and 2 are presented in Fig. 4g. **i**, Mean and s.d. of the percentage of mature adipocytes developed within the plugs composed of either Lin^-SCA1^+ or $Lin^-SCA1^+CD142^-$ ASPCs ($n = 7$ biological replicates) analysed using CellProfiler[25]. $^*P \leq 0.01$, paired *t*-test. Mean and s.d. are shown. **j**, Microscopy images of Matrigel plugs corresponding to **i**. Scale bar, 100 μm. **k**, Microscopy images of Matrigel plugs stained with isolectin GS-IB4 (green) and Hoechst (blue) corresponding to **i**. Scale bar, 100 μm. **l**, Quantification of vascularisation within the plugs, composed of either Lin^-SCA1^+ or $Lin^-SCA1^+CD142^-$ ASPCs ($n = 6$ biological replicates) analysed using CellProfiler[25].

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 8 | Supplementary methods for the scRNA-seq analyses. a, **b**, *t*-SNE 2D maps of all analysed cells (Fluidigm C1), highlighting the four subpopulations (P1–P4) obtained when clustering analysis was performed with $k = 4$ (**a**) and cells belonging to one of three biological replicates (**b**) (pink, R1; green, R2; blue, R3). **c**, For each of the four clusters shown in **a**, the number of cells stemming from each biological replicate is highlighted. **d**, Silhouette analysis results for $k = 3$ and $k = 4$. **e**, Venn diagram showing the overlap between the 527 significantly differentially expressed genes identified using M3Drop in the Letter, and the 1,827 genes showing high variability (HVG). **f**, For each cluster shown in Fig. 1b, we determined the number of cells contained in the alternative clustering obtained by considering 1,827 highly variable genes for SC3 analysis. The vast majority of cells are similarly attributed, irrespective of gene set choice. **g**, Percentage of the top 100 marker genes of the three subpopulations described in the Letter that overlap with the top 100 marker genes of the subpopulations derived from the analysis based on highly variable genes. **h**, *t*-SNE 2D maps of all analysed cells (C1) considering only the highly variable genes, highlighting the three subpopulations (P1–P3) identified by cluster analysis in the Letter (M3Drop, right) and using the highly variable genes (HVG, left). In **a**, **b**, **h**, $n = 208$ cells from three independent biological experiments, each based on a pool of mice. **i**, Marker gene expression across all 10x Genomics cells, including a housekeeping gene (*Actb*), the negative FACS markers *Ptprc* (*Cd45*) and *Pecam1* (*Cd31*), the positive FACS markers *Cd34*, *Ly6a* (*Sca1*) and *Itgb1* (*Cd29*), and the mature adipogenic markers *Adipoq*, *Retn* and *Cidec*. **j**, *t*-SNE 2D maps of all 10x Genomics cells highlighting the expression of *Xist* and the cell division marker *Mki67* (black). **k**, Sum of squares for distinct number of clusters (2 to 15) in the 10x Genomics data. In **i**, **j**, $n = 2{,}919$ cells from one biological experiment based on a pool of mice, before G1, *Xist*, *Krt18* or *Krt19* and *Epcam* filtering. **l**, Silhouette width for distinct number of clusters (2 to 13) in the 10x Genomics data. **m**, Percentage of the top 100 marker genes of the 10x Genomics populations (G1–G4) that overlaps with the top 100 marker genes of the populations (P1–P3) derived from the C1 data based on the highly variable genes.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

https://doi.org/10.1038/s41586-018-0257-1

Parasitic helminths induce fetal-like reversion in the intestinal stem cell niche

Ysbrand M. Nusse[1,2,6], Adam K. Savage[3,6], Pauline Marangoni[2], Axel K. M. Rosendahl-Huber[2], Tyler A. Landman[2], Frederic J. de Sauvage[4], Richard M. Locksley[3,6]* & Ophir D. Klein[2,5,6]*

Epithelial surfaces form critical barriers to the outside world and are continuously renewed by adult stem cells[1]. Whereas dynamics of epithelial stem cells during homeostasis are increasingly well understood, how stem cells are redirected from a tissue-maintenance program to initiate repair after injury remains unclear. Here we examined infection by *Heligmosomoides polygyrus*, a co-evolved pathosymbiont of mice, to assess the epithelial response to disruption of the mucosal barrier. *H. polygyrus* disrupts tissue integrity by penetrating the duodenal mucosa, where it develops while surrounded by a multicellular granulomatous infiltrate[2]. Crypts overlying larvae-associated granulomas did not express intestinal stem cell markers, including *Lgr5*[3], in spite of continued epithelial proliferation. Granuloma-associated Lgr5$^-$ crypt epithelium activated an interferon-gamma (IFN-γ)-dependent transcriptional program, highlighted by Sca-1 expression, and IFN-γ-producing immune cells were found in granulomas. A similar epithelial response accompanied systemic activation of immune cells, intestinal irradiation, or ablation of Lgr5$^+$ intestinal stem cells. When cultured in vitro, granuloma-associated crypt cells formed spheroids similar to those formed by fetal epithelium, and a sub-population of *H. polygyrus*-induced cells activated a fetal-like transcriptional program, demonstrating that adult intestinal tissues can repurpose aspects of fetal development. Therefore, re-initiation of the developmental program represents a fundamental mechanism by which the intestinal crypt can remodel itself to sustain function after injury.

To study how intestinal crypts respond to tissue disruption, we infected *Lgr5*$^{DTRGFP/+}$ (Lgr5–GFP) reporter mice[4] with *H. polygyrus*. Six days after infection, larvae resided within the intestinal wall, and were surrounded by an immune infiltrate. Crypts overlying granulomas

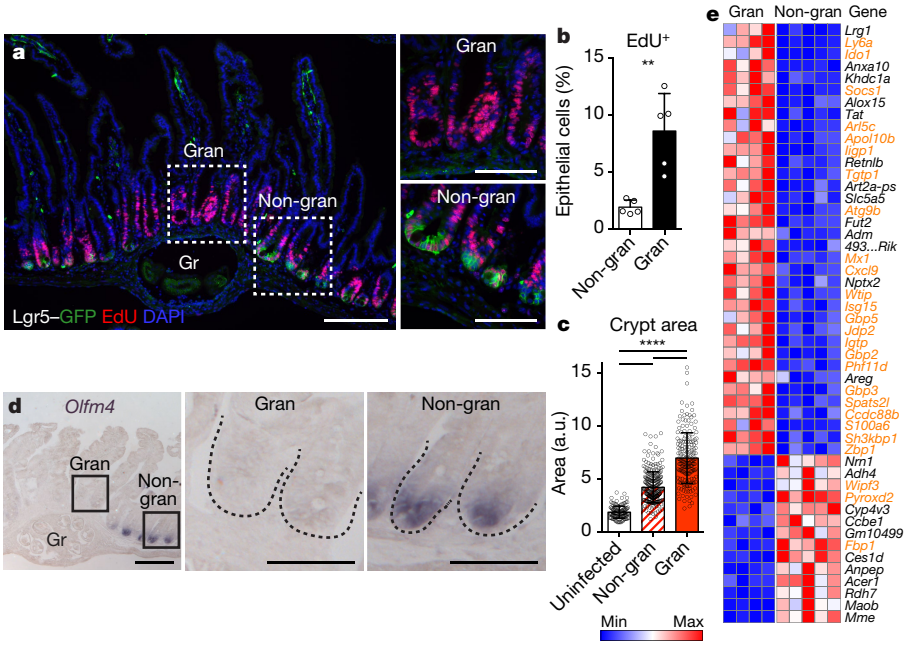


Fig. 1 | Helminth infection induces an Lgr5$^-$ program in affected crypt epithelium. Analysis of crypts overlying (gran) or adjacent to (non-gran) *H. polygyrus* granulomas (Gr) from day 6 of infection. **a**, Representative image of Lgr5–GFP and EdU staining in thin sections. **b**, Flow cytometry of EdU in biopsies of total epithelium. **c**, Crypt area from uninfected or infected (gran or non-gran) mice; a.u., arbitrary units. **d**, Representative image of in situ hybridization for *Olfm4* in thin sections. **e**, RNA-seq of sorted crypt epithelium from non-granuloma or granuloma biopsies. Data were filtered for ≥100 reads average in either group, FDR ≤ 10^{-4}, and the 50 genes with the greatest fold-change are presented; high (red) and low (blue) relative expression. Gene names in orange are predicted IFN targets. *493…Rik*, RIKEN cDNA 4930461G14 gene (*4930461G14Rik*). $n = 5$ mice (**a**, **b**, **d**); 6 mice (**c**, uninfected); 15 mice (**c**, infected); 4 independently sorted samples (**e**, granuloma, 20 mice total), or 5 independently sorted samples (**e**, non-granuloma, 25 mice total). Statistics represent all biological replicates, and all experiments were replicated at least twice. Graphs show mean ± s.d. (**b**, **c**). $^{**}P < 0.01$, $^{****}P < 0.0001$ by unpaired, two-tailed Mann–Whitney test. Scale bars: main, 200 μm (**a** left, **d** left), 100 μm (magnified insets in **a**), 50 μm (insets in **d**).

[1]Biomedical Sciences Graduate Program, University of California, San Francisco, CA, USA. [2]Program in Craniofacial Biology and Department of Orofacial Sciences, University of California, San Francisco, CA, USA. [3]Howard Hughes Medical Institute and Departments of Medicine and Microbiology & Immunology, University of California, San Francisco, CA, USA. [4]Department of Molecular Oncology, Genentech Inc., South San Francisco, CA, USA. [5]Department of Pediatrics and Institute for Human Genetics, University of California, San Francisco, CA, USA. [6]These authors contributed equally: Ysbrand M. Nusse, Adam K. Savage, Richard M. Locksley, Ophir D. Klein. *e-mail: richard.locksley@ucsf.edu; ophir.klein@ucsf.edu

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

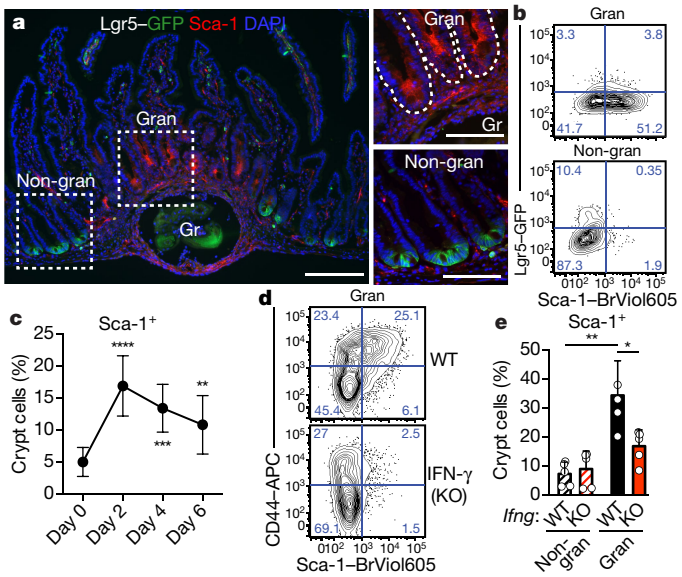


Fig. 2 | IFN-γ mediates the helminth-induced crypt phenotype. Day 6 of *H. polygyrus* infection, except as noted. **a**, Representative image of Lgr5–GFP and Sca-1 staining in crypts overlying (gran) or adjacent to (non-gran) granulomas (Gr). **b**, Representative flow cytometry of Lgr5–GFP and Sca-1 in crypt biopsies. **c**, Flow cytometry of Sca-1 in crypts from unfractionated epithelium following infection. **d**, Representative flow cytometry of CD44 and Sca-1 in epithelia from biopsies of IFN-γ-knockout (KO) mice. APC, allophycocyanin. **e**, Cells analysed as in (**d**). $n = 4$ (**b**), 5 (**a**, **d**, **e**), 8 (**c**, day 2, 4, 6), or 9 mice (**c**, day 0). Statistics represent all biological replicates, and all experiments were replicated at least twice. Graphs show mean ± s.d. (**c**, **e**). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$ by unpaired, two-tailed Mann–Whitney test. Scale bars: 200 μm (**a**, left), 100 μm (**a**, magnified insets).

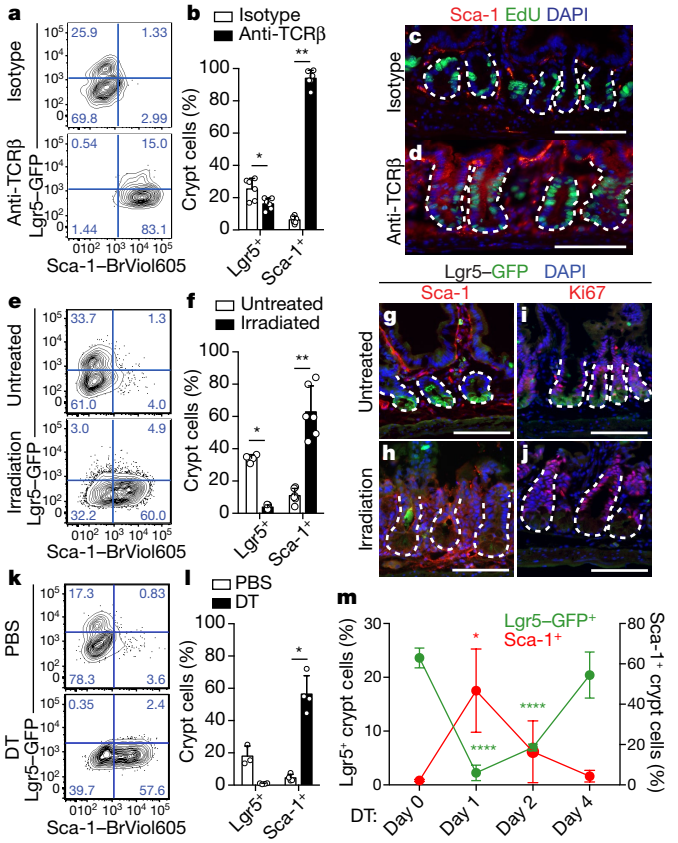


Fig. 3 | The crypt response to *H. polygyrus* is a generalized response to tissue injury. a–d, Mice were treated with 20 μg anti-TCRβ and analysed 24 h later by flow cytometry for Lgr5–GFP and Sca-1 in crypt cells (**a**, **b**) or for Sca-1 and EdU in thin sections (**c**, **d**). **e–j**, Mice were treated with 10 Gy irradiation and analysed at three days by flow cytometry and in thin sections for Lgr5–GFP and Sca-1 in crypt cells (**e–h**) or in thin sections for Lgr5–GFP and Ki67 (**i**, **j**). **k–m**, *Lgr5*$^{DTRGFP/+}$ mice were treated with diphtheria toxin (DT) and analysed by flow cytometry for Lgr5–GFP and Sca-1 in crypt cells at 24 h (**k**, **l**) or indicated time points (**m**). $n = 3$ (**g–l**, Lgr5, untreated; **m**, day 0), 4 (**c–f**, Lgr5, untreated; **k**, **l**, all others; **m**, day 1, 2, 4), 5 (**e**, **f**, Lgr5, irradiated), or 6 mice (**a**, **b**, **f**, Sca-1, irradiated). Statistics represent all biological replicates, and all experiments were replicated at least twice. Graphs show mean ± s.d. (**b**, **f**, **l**, **m**). $^*P < 0.05$, $^{**}P < 0.01$, $^{****}P < 0.0001$ by unpaired, two-tailed Mann–Whitney test (**b**, **f**, **l**) or unpaired, two-tailed *t*-tests (**m**). Scale bars: 100 μm (**c**, **d**, **g–j**).

(granuloma-associated crypts or GACs) were hyper-proliferative and enlarged (Fig. 1a–c, Extended Data Fig. 1a), as previously reported[5]. Notably, GACs did not express the Lgr5–GFP reporter (Fig. 1a, Extended Data Fig. 1b), whereas crypts not associated with granulomas still expressed Lgr5–GFP (Fig. 1a). *Olfm4*, another marker of intestinal stem cells (ISCs), was similarly repressed (Fig. 1d). In addition to loss of *Lgr5* and *Olfm4* expression, the Paneth cell marker MMP7 frequently co-stained with the goblet cell marker MUC2 (Extended Data Fig. 1c, d), as previously observed in helminth infections[6] and other perturbations of epithelial lineage commitment[7]. Thus, the epithelium overlying granulomas exhibits loss of ISC markers and disruption of the ISC niche[8].

To identify pathways in GACs that respond to *H. polygyrus* infection, we performed RNA sequencing analysis (RNA-seq) on purified crypt epithelium from punch biopsies of granulomas (Extended Data Fig. 2a). We identified 277 genes that were differentially expressed between biopsies of granuloma and non-granuloma crypts (Fig. 1e, Extended Data Fig. 2b, Supplementary Table 1). In addition to *Lgr5* and *Olfm4*, a suite of ISC signature genes[9] was downregulated in GACs (Extended Data Fig. 2c, d, Supplementary Table 2), confirming that *H. polygyrus* infection represses ISCs. Among the genes that were upregulated in GACs were an abundance of targets of interferon (IFN) signalling (Fig. 1e), and pathway analysis revealed an IFN response (Extended Data Fig. 2e, Supplementary Table 2).

One of the most highly upregulated genes was *Ly6a*, which encodes Sca-1, a surface protein that is associated with proliferative cells but is not present in humans. Sca-1 is a target of IFN signalling, and is induced in epithelia during colitis[10]. Immunofluorescence analysis showed that Sca-1 specifically marked Lgr5–GFP$^-$ GACs (Fig. 2a). Flow cytometry analysis confirmed that Sca-1 was enriched in GAC biopsies (Fig. 2b) and revealed that Sca-1 upregulation occurred as early as two days after infection (Fig. 2c, Extended Data Fig. 3a). Furthermore, Sca-1 expression was inversely correlated with that of Lgr5–GFP at all time points. Therefore, Sca-1 was a useful marker of crypt cells responding to *H. polygyrus*-driven epithelial disruption. By day 10 post-infection, diminished Sca-1 expression at granuloma remnants (Extended Data Fig. 3b) indicated that resolution had commenced. Another intestinal helminth, *Nippostrongylus brasiliensis*, which does not invade intestinal tissue, did not induce Sca-1 expression (Extended Data Fig. 3c), suggesting that Sca-1 expression is a specific response to crypt disruption.

Although helminths are typically associated with allergic immunity[11], our data pointed towards a role for IFN. We focused on IFN-γ, because elevated expression of this gene was found in granulomas of infected mice (Extended Data Fig. 3d), whereas type I and type III IFN transcripts were not induced in GACs (Extended Data Fig. 3e). We also found large numbers of neutrophils, which are known targets of IFN-γ[12], and an accumulation of IFN-γ$^+$ lymphocytes in granulomas (Extended Data Fig. 4a–d). Infection of IFN-γ-null mice with *H. polygyrus* showed that induction of Sca-1 (Fig. 2d, e) and IFN-target genes (Extended Data Fig. 4e) was dependent on IFN-γ; however, the Lgr5–GFP reporter remained downregulated in GACs of IFN-γ-null mice (Extended Data Fig. 4f). To assess the cell-autonomous effects of IFN-γ on intestinal epithelia, we deleted the IFN-γ receptor in intestinal epithelium and found a similar effect as with germline deletion of IFN-γ (Extended Data

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

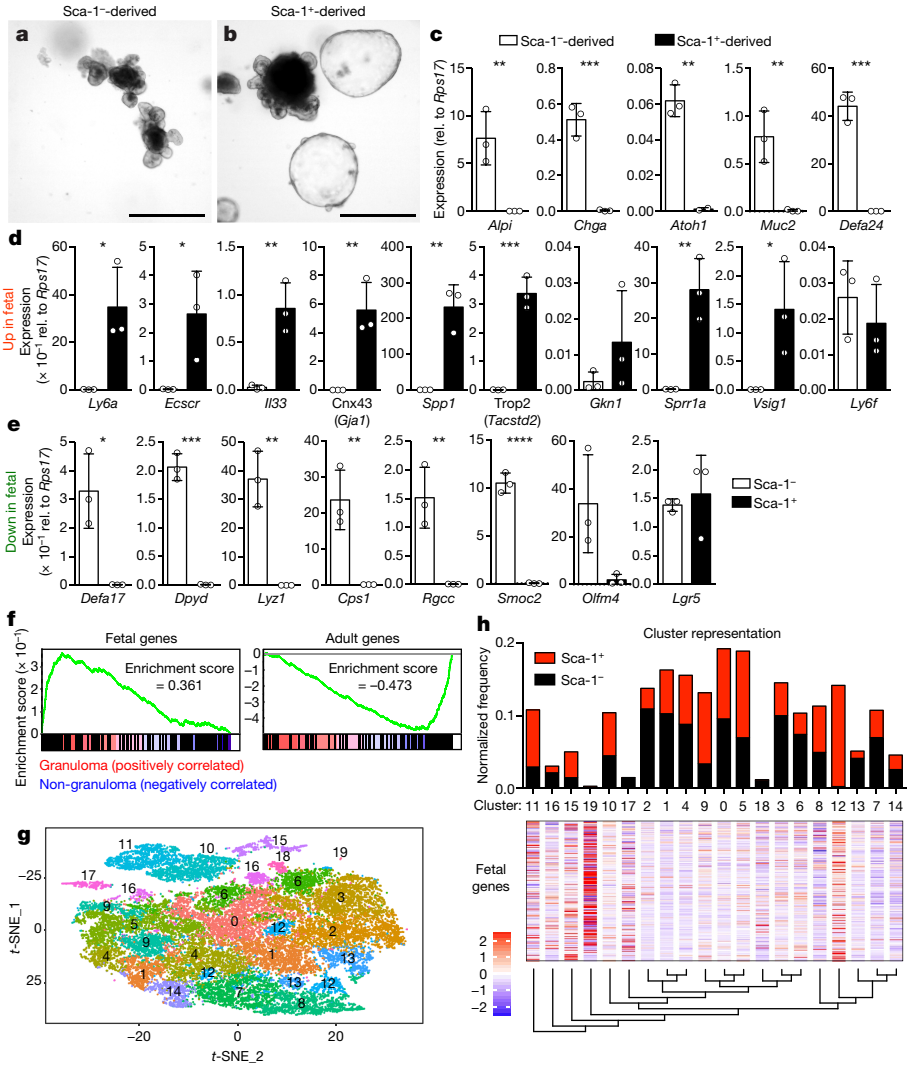


Fig. 4 | Helminth-associated crypts acquire a fetal-like program. **a**–**e**, Sorted Sca-1$^-$ (**a**) or Sca-1$^+$ (**b**) crypt cells from *H. polygyrus*-infected mice were cultured in organoid conditions, imaged after one passage (**a**, **b**), and analysed by quantitative PCR (qPCR) for markers of differentiated cells (**c**) or fetal-derived cultures[19] (**d**, **e**). **f**, Bulk RNA-seq data (as in Fig. 1e) were analysed by gene set enrichment analysis (GSEA) for cell signature genes[19]. All analyses have false discovery rate (FDR) $<10^{-3}$. **g**, **h**, Single-cell RNA-seq from $n = 19{,}754$ Sca-1$^-$ and $n = 6{,}669$ Sca-1$^+$ individually sorted crypt cells from one *H. polygyrus*-infected mouse. **g**, *t*-distributed stochastic neighbour embedding (*t*-SNE) distribution, colour coded to represent clusters identified independently by unsupervised hierarchical clustering. The relation of cluster identity to transcriptional signatures of mature lineages is shown in Extended Data Fig. 9. **h**, Sca-1$^-$ and Sca-1$^+$ cell frequency within each cluster, normalized to the total number of cells sequenced from each population (top). Normalized expression values for the fetal gene signature[19] were mapped to the clusters (middle) and arranged per the unsupervised dendrogram of cluster relatedness (bottom). $n = 3$ cultures from 3 mice (**c**–**e**), 4 independently sorted samples (**f**, granuloma, 20 mice total), 5 independently sorted samples (**f**, non-granuloma, 25 mice total), or 15 cultures from 15 mice (**a**, **b**). Statistics represent all biological replicates, and all experiments were replicated at least twice, except the single cell experiment, which was performed once. Graphs show mean ± s.d. (**c**–**e**). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$ by unpaired two-tailed *t*-tests. Scale bars: 500 μm (**a**, **b**).

Fig. 4g). Treating intestinal organoids with IFN-γ led to transcriptional changes corresponding to those found in GACs (Extended Data Fig. 4h). Together, these data demonstrate that immune cell-derived IFN-γ is a critical component of the GAC response.

Lymphocyte activation and IFN-γ production are elicited in other contexts of epithelial injury[13–15]. Therefore, we challenged uninfected mice with a monoclonal antibody (anti-TCRβ) that cross-links and activates the T cell receptor to assess the host response to immune cell activation. After 24 h, the level of *Ifng* transcript was increased (Extended Data Fig. 5a), and the intestinal epithelium broadly resembled the *H. polygyrus* GAC response, as indicated by reduction of Lgr5–GFP expression, induction of Sca-1, increased proliferation, increased crypt size (Fig. 3a–d, Extended Data Fig. 5b, c), and expression of a subset of *H. polygyrus*-activated transcriptional targets (Extended Data Fig. 5d).

The convergence of epithelial responses to immune cell activation following *H. polygyrus* infection and anti-TCRβ challenge might reflect a generalized reaction to tissue perturbation. To test this, we examined other tissue injury models. First, we lethally irradiated mice and analysed them after three days, at which time *Lgr5* expression is lost during regeneration[16,17]. We observed Sca-1 induction and an IFN response in crypt cells (Fig. 3e–h, Extended Data Fig. 5e, f), as well as continued proliferation (Fig. 3i, j) and increased crypt depth (Extended Data Fig. 5g), as previously reported[16].

Because irradiation is relatively non-specific, we sought to restrict cell death to the stem cell compartment by specifically ablating *Lgr5*-expressing cells[4]. Twenty-four hours after treatment of *Lgr5*$^{DTRGFP/+}$ mice with diphtheria toxin, Lgr5–GFP$^+$ cells were absent and Sca-1 was highly induced (Fig. 3k, l, Extended Data Fig. 6a, b). During recovery, Lgr5–GFP$^+$ ISCs re-emerged and Sca-1 expression decreased to

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

baseline levels (Fig. 3m). Notably, ablation of Lgr5$^+$ cells did not induce crypt hyperplasia (Extended Data Fig. 6c) or GAC-like expression of some IFN targets (data not shown), showing that there is a distinction between the Sca-1 response following *Lgr5* ablation and the response to other types of epithelial disruption. However, IFN activation has been observed after ablation of Lgr5$^+$ cells in tumours[18]. Therefore, diverse insults that disrupt Lgr5$^+$ cells induce GAC-like responses during regeneration.

Lgr5$^+$ cells are required for regeneration after irradiation-induced injury[17]. To test whether they are also required for the *H. polygyrus*-induced GAC phenotype, we ablated Lgr5$^+$ cells immediately before infection. In this setting, although we confirmed that GAC cells were part of the Lgr5$^+$ ISC lineage hierarchy (Extended Data Fig. 6d, e), crypt cell frequency, Sca-1 induction, and 5-ethynyl-2′-deoxyuridine (EdU) incorporation were unaffected (Extended Data Fig. 6f–k). These data indicate that whereas the *H. polygyrus*-induced GAC phenotype is mediated by progeny of ISCs, it can occur independently of Lgr5$^+$ ISCs.

Sca-1$^+$ GAC cells were hyper-proliferative and gave rise to granuloma-associated villus epithelium (Extended Data Fig. 7a, b). To assess the regenerative capacity of GACs, we sorted Sca-1$^+$ and Sca-1$^-$ crypt cells from *H. polygyrus*-infected mice and cultured them under standard organoid conditions. Whereas Sca-1$^-$ cells formed typical organoids (Fig. 4a), Sca-1$^+$ cells formed large, smooth spheroids that were devoid of crypt budding (Fig. 4b) and could be stably passaged for more than six months (data not shown). Sca-1$^+$ spheroids lost expression of markers of differentiated epithelium (Fig. 4c), suggesting that they reflected growth of an undifferentiated cell type. Spheroids have been observed in high-Wnt conditions[8]; however, in our studies, we did not add exogenous Wnt, and we found no difference in *Axin2* expression between Sca-1$^+$ and Sca-1$^-$ cultures (Extended Data Fig. 8a), suggesting that Wnt signalling is not hyperactive in Sca-1$^+$ spheroids. Recent work[19,20] has demonstrated that cultured fetal epithelium also forms spheroids. We therefore tested expression of fetal epithelial markers and found that nearly all fetal genes assayed were highly expressed in Sca-1$^+$ cultures (Fig. 4d, e). Like fetal cultures[20], Sca-1$^+$ spheroids were not sensitive to R-Spondin1 withdrawal (data not shown). Therefore, Sca-1$^+$ cells adopted a state in vitro that was distinct from Sca-1$^-$ cells and exhibited characteristics of fetal intestinal epithelium.

We sought to determine whether the fetal program was activated in vivo and found that the fetal markers *Gja1* and *Spp1* were upregulated in GACs during *H. polygyrus* infection (Extended Data Fig. 8b). Furthermore, Sca-1 was expressed in mouse fetal intestinal epithelium at embryonic day 15.5 (Extended Data Fig. 8c). This remarkable similarity led us to re-analyse the results of our RNA-seq of GAC epithelium. We found strong enrichment of the fetal signature in GAC epithelium, whereas the adult signature was enriched in non-GACs (Fig. 4f, Supplementary Table 2). Furthermore, the signatures[21] of stem cells, enterocytes and Paneth cells, but not goblet cells, were lost in GACs (Extended Data Fig. 8d, Supplementary Table 2). Taken together, these data indicate that, during infection, GACs adopt an undifferentiated state resembling the fetal epithelium.

Enrichment of the goblet cell signature in GAC epithelium suggests that there is heterogeneity within the pool of Sca-1$^+$ cells. To investigate whether a subgroup of cells underpin the fetal signature, we performed single-cell RNA-seq of Sca-1$^+$ and Sca-1$^-$ crypt cells. Unsupervised clustering of the merged datasets revealed that most cell clusters were composed of both Sca-1$^+$ and Sca-1$^-$ cells (Fig. 4g, h) and, although we excluded mature epithelium, the transcriptional signature of specific lineages could be recognized in some clusters (Extended Data Fig. 9, Supplementary Table 4). We focused on cluster 12, which consisted almost entirely of Sca-1$^+$ cells (98.2%). By overlaying known intestinal cell type signatures[19,21], we found that cluster 12 was depleted of mature cell markers (Extended Data Fig. 9), and was strongly enriched for the fetal program (Fig. 4h, Supplementary Table 5), suggesting that this cluster represents a unique cell identity within the larger Sca-1$^+$ pool that is elicited by *H. polygyrus* infection.

Intestinal crypts have been postulated to respond to damage by activation of reserve stem cells[16,22] or reacquisition of stemness by differentiated progenitors[23–25]. Here, our use of an evolutionarily adapted parasite led to identification of a novel infection-mediated alteration of the crypt in response to injury. By monitoring the markers Lgr5–GFP and Sca-1, we found that an overlapping injury-response program was induced by other tissue-damaging agents, indicating a generalized strategy by which the intestine responds to stress. Our data identify a novel cell type that arises in the damaged crypts and suggest that crypt repair repurposes aspects of fetal development in order to restore integrity of the intestinal barrier. Indeed, a re-activation of fetal markers has been observed in models of injury in other tissues[26–28]. Furthermore, a recent report examining a chemical colitis model also uncovered induction of Sca-1 and a fetal signature in regenerative colonic crypts[29]. This work showed that extracellular matrix-driven activation of Yap–Taz signalling was linked to regeneration, whereas we found a discrete subset of Sca-1$^+$ crypt cells associated with a fetal gene signature. These complementary studies suggest that induction of the fetal program may be a core response to injury in the crypt. Taken together, the helminth-induced changes in crypt epithelia that we discovered point to a repurposing of some of the functional capabilities of the developing fetal gut and highlight a novel mechanism of repair in the intestinal crypt that involves infection-induced developmental plasticity.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0257-1.

Received: 13 April 2018; Accepted: 9 May 2018;
Published online: 27 June 2018

1. Karin, M. & Clevers, H. Reparative inflammation takes charge of tissue regeneration. *Nature* **529**, 307–315 (2016).
2. Maizels, R. M. et al. Immune modulation and modulators in *Heligmosomoides polygyrus* infection. *Exp. Parasitol.* **132**, 76–89 (2012).
3. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
4. Tian, H. et al. A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. *Nature* **478**, 255–259 (2011).
5. Ferguson, A. & Jarrett, E. E. Hypersensitivity reactions in small intestine. I Thymus dependence of experimental 'partial villous atrophy'. *Gut* **16**, 114–117 (1975).
6. Kamal, M. et al. Paneth and intermediate cell hyperplasia induced in mice by helminth infections. *Parasitology* **125**, 275–281 (2002).
7. VanDussen, K. L. et al. Notch signaling modulates proliferation and differentiation of intestinal crypt base columnar stem cells. *Development* **139**, 488–497 (2012).
8. Sato, T. et al. Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011).
9. Muñoz, J. et al. The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. *EMBO J.* **31**, 3079–3091 (2012).
10. Flanagan, K. et al. Intestinal epithelial cell up-regulation of LY6 molecules during colitis results in enhanced chemokine secretion. *J. Immunol.* **180**, 3874–3881 (2008).
11. Grencis, R. K. Immunity to helminths: Resistance, regulation, and susceptibility to gastrointestinal nematodes. *Annu. Rev. Immunol.* **33**, 201–225 (2015).
12. Amulic, B., Cazalet, C., Hayes, G. L., Metzler, K. D. & Zychlinsky, A. Neutrophil function: From mechanisms to disease. *Annu. Rev. Immunol.* **30**, 459–489 (2012).
13. Miura, N. et al. Anti-CD3 induces bi-phasic apoptosis in murine intestinal epithelial cells: possible involvement of the Fas/Fas ligand system in different T cell compartments. *Int. Immunol.* **17**, 513–522 (2005).
14. Sollid, L. M. & Jabri, B. Triggers and drivers of autoimmunity: lessons from coeliac disease. *Nat. Rev. Immunol.* **13**, 294–302 (2013).
15. Zhou, P., Streutker, C., Borojevic, R., Wang, Y. & Croitoru, K. IL-10 modulates intestinal damage and epithelial cell apoptosis in T cell-mediated enteropathy. *Am. J. Physiol. Gastrointest. Liver Physiol.* **287**, G599–G604 (2004).
16. Yan, K. S. et al. The intestinal stem cell markers Bmi1 and Lgr5 identify two functionally distinct populations. *Proc. Natl Acad. Sci. USA* **109**, 466–471 (2011).
17. Metcalfe, C., Kljavin, N. M., Ybarra, R. & de Sauvage, F. J. Lgr5$^+$ stem cells are indispensable for radiation-induced intestinal regeneration. *Cell Stem Cell* **14**, 149–159 (2014).
18. Melo, F. de S. E. et al. A distinct role for Lgr5$^+$stem cells in primary and metastatic colon cancer. *Nature* **543**, 676–680 (2017).
19. Mustata, R. C. et al. Identification of Lgr5-independent spheroid-generating progenitors of the mouse fetal intestinal epithelium. *Cell Reports* **5**, 421–432 (2013).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

20. Fordham, R. P. et al. Transplantation of expanded fetal intestinal progenitors contributes to colon regeneration after injury. *Cell Stem Cell* **13**, 734–744 (2013).
21. Haber, A. L. et al. A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).
22. Asfaha, S. et al. $Krt19^{+}/Lgr5^{-}$ cells are radioresistant cancer-initiating stem cells in the colon and intestine. *Cell Stem Cell* **16**, 627–638 (2015).
23. van Es, J. H. et al. $Dll1^{+}$ secretory progenitor cells revert to stem cells upon crypt damage. *Nat. Cell Biol.* **14**, 1099–1104 (2012).
24. Tetteh, P. W. et al. Replacement of lost Lgr5-positive stem cells through plasticity of their enterocyte-lineage daughters. *Cell Stem Cell* **18**, 203–213 (2016).
25. Buczacki, S. J. A. et al. Intestinal label-retaining cells are secretory precursors expressing Lgr5. *Nature* **495**, 65–69 (2013).
26. Fernandez Vallone, V. et al. Trop2 marks transient gastric fetal epithelium and adult regenerating cells after epithelial damage. *Development* **143**, 1452–1463 (2016).
27. Gadye, L. et al. Injury activates transient olfactory stem cell states with diverse lineage capacities. *Cell Stem Cell* **21**, 775–790 (2017).
28. Lin, B. et al. Injury induces endogenous reprogramming and dedifferentiation of neuronal progenitors to multipotency. *Cell Stem Cell* **21**, 761–774 (2017).
29. Yui, S. et al. Yap/Taz-dependent reprogramming of colonic epithelium links ECM remodeling to tissue regeneration. *Cell Stem Cell* **22**, 35–49 (2018).

Acknowledgements We thank M. Consengco, R. D'Urso, J. Ming, A. Rathnayake, N. Wang and Z. Wang for technical expertise, the UCSF Institute of Human Genetics Core and Functional Genomics Core for performing the RNA-seq experiments, members of the Klein and Locksley labs for discussions, and K. Lindquist for advice on GSEA analysis. This work was supported by the National Institutes of Health (AI026918, AI030663 and U01DK103147 from the Intestinal Stem Cell Consortium—a collaborative research project funded by the National Institute of Diabetes and Digestive and Kidney Diseases and the National Institute of Allergy and Infectious Diseases), the Howard Hughes Medical Institute (HHMI), the California Institute for Regenerative Medicine (RN3-06525) and the Sandler Asthma Basic Research Center at the University of California, San Francisco. A.K.S. is an HHMI Fellow. Y.M.N. was awarded a Genentech Graduate Fellowship in 2014.

Reviewer information *Nature* thanks A. Hanash, T. Sato, S. Takashima and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions Y.M.N. and A.K.S. jointly conceived the study and designed, performed and interpreted histology, cytometry, transcriptomic and culture experiments. P.M. performed computational analysis of the single-cell transcriptomic experiment. A.K.M.R.-H. and T.A.L. performed and interpreted histology experiments. F.J.d.S. contributed essential reagents. O.D.K. and R.M.L. directed the study and wrote the manuscript with Y.M.N. and A.K.S.

Competing interests F.J.d.S. is an employee of Genentech and owns shares in Roche.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0257-1.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0257-1.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to R.M.L. and O.D.K.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Mice. Mice were maintained in the University of California San Francisco (UCSF) specific pathogen-free animal facility in compliance with all ethical guidelines established by the Institutional Animal Care and Use Committee and Laboratory Animal Resource Center. All experimental procedures were approved by the Laboratory Animal Resource Center at UCSF. Both male and female mice aged 6–14 weeks were used for all experiments, except those analysing fetal tissue. Mice were chosen for experimental conditions randomly but without formal randomization. Investigators were blinded where possible, but this was frequently precluded owing to inherent facets of infection experiments. *Lgr5*^{DTRGFP} mice were previously described[4]. Wild-type (C57BL/6J), *Lgr5*$^{GFP\text{-}creERT2/+}$ (B6.129P2-*Lgr5*$^{tm1(cre/ERT2)Cle}$/J), *Rosa26*$^{RFP/+}$ (B6;129S6-*Gt(ROSA)26Sor*$^{tm14(CAG\text{-}tdTomato)Hze}$/J), IFN-γ reporter (B6.129S4-*Ifng*$^{tm3.1Lky}$/J), IFN-γ-null (B6.129S7-*Ifng*^{tm1Ts}/J), IFN-γ receptor-flox (*Ifngr1*$^{loxP/loxP}$; C57BL/6N-*Ifngr1*$^{tm1.1Rds}$/J), and Vil1-cre (B6.Cg-Tg(Vil1-cre)997Gum/J) mice were from The Jackson Laboratory and were maintained on a C57BL/6 background. For analysis of embryonic tissue, timed matings were established, and the morning after plugs were recognized was considered embryonic stage (e) 0.5. Fetal intestine was dissected at the time points indicated.

Helminth infection and treatments. Mice were infected by oral gavage with 200 *H. polygyrus* L3 larvae and were killed at the indicated time points. For anti-TCRβ treatment, mice were administered with 20 μg per mouse of clone H57 TCRβ monoclonal antibody and analysed 24 h later. For irradiation, mice were exposed to 10 Gy and analysed approximately 72 h later. For ablation of Lgr5^{+} cells, *Lgr5*$^{DTRGFP/+}$ mice were administered 50 μg/kg diphtheria toxin intraperitoneally and analysed at the indicated time points. For determination of cell proliferation, 500 μg EdU was administered intraperitoneally 1 h before mice were killed, except for Extended Data Fig. 7b, where EdU was administered 24 h before mice were killed. For lineage tracing experiments, *Lgr5*$^{GFP\text{-}creERT2/+}$ *Rosa26*$^{RFP/+}$ mice were infected with *H. polygyrus* and injected with 2.5 mg tamoxifen in corn oil intraperitoneally at the indicated time points and analysed at day 6.

Tissue preparation and flow cytometry. Preparation of intestinal tissue for flow cytometry was modified from previous work[30]. The duodenum was dissected, flushed extensively with cold PBS, and the mesenteric tissue was removed. For whole tissue preps, Peyer's patches were removed and tissue was turned inside out. For recovery of punch biopsies, tissue was filleted open longitudinally. In both cases, tissue was shaken in three changes of 20 ml cold PBS and washed for 20 min at 37 °C in two changes of 20 ml Ca^{2+}/Mg^{2+}-free HBSS containing 5 mM DTT, 10 mM HEPES and 2% FCS, followed by 20 ml of Ca^{2+}/Mg^{2+}-replete HBSS containing 10 mM HEPES and 2% FCS. For punch biopsies, granuloma and non-granuloma tissues were dissected with a 1-mm punch tool under low-power magnification. Tissues were digested for 30 min at 37 °C in 5 ml (whole tissue) or 2 ml (punch biopsies) Ca^{2+}/Mg^{2+}-replete HBSS containing 10 mM HEPES, 2% FCS, 30 μg/ml DNaseI (Roche) and 0.1 Wünsch units/ml Liberase TM (Roche), and whole tissue was homogenized in C tubes using a gentleMACS tissue dissociator (Miltenyi). Homogenate or punch biopsies were passed through a 100-μm filter with assistance of a 3-ml syringe plunger, and enumerated for staining equivalent numbers for flow cytometry or sorting. Fc Block (anti-CD16/32), doublet exclusion, and DAPI exclusion were used in all cases. Data were acquired with a Becton Dickinson Fortessa and analysed using FlowJo (Tree Star). Cell sorting was performed with a Beckman Coulter MoFlo XDP.

Immunofluorescence and in situ hybridization. For tissue staining in section, mice were perfused with cold 4% PFA in PBS. The proximal 10 cm of duodenum was cleaned, flushed with cold 4% PFA, and fixed in 4% PFA for 4 h at 4 °C. The tissue was cryo-protected in 30% sucrose overnight at 4 °C. Samples were embedded in OCT and 8-μm sections were prepared for immunofluorescence staining. For immunohistochemistry and in situ hybridization, the samples were fixed overnight in 4% PFA, paraffin embedded, and sectioned at 5 μm for immunohistochemistry or 10 μm for in situ hybridization. For crypt area quantification, crypts clearly above distended granuloma tissue containing visible larval worms were called as 'granuloma' and others were called 'non-granuloma'. Crypt area was quantified in ImageJ. For fetal whole mount imaging, fetal intestines were fixed in 4% PFA in PBS for 3 h, permeabilized, and blocked for 4 h at room temperature. Primary and secondary antibodies were incubated at 4 °C overnight. Images were acquired and processed with a Leica DM5000 B or a Zeiss Axio Imager 2 and Adobe Photoshop.

In situ hybridization. An *Olfm4* probe was designed by PCR amplifying an 898-bp sequence from total intestinal cDNA using the primers 5′-AACCTG ACGGTCCGAGTAGA-3′ (forward) and 5′-TGCTGGCCTCAGTTGCATAA-3′ (reverse). *Olfm4* cDNA was cloned into a pGEM-T Easy vector (Promega). *Olfm4* antisense probes were prepared and in situ hybridization was performed as described[31].

Bulk RNA sequencing. Five wild-type mice were infected with *H. polygyrus* and six days later 1-mm punch biopsies from granuloma and non-granuloma tissue were taken, pooled by tissue, digested, and sorted for DAPI^{lo} CD45^{-} EpCAM^{+} CD44^{+} crypt epithelium. RNA from six granuloma (30 mice total) and five non-granuloma (25 mice total) sorts was submitted for RNA sequencing. Two granuloma datasets were excluded owing to low unique mapping rates and failure to group by tissue in principle components analysis and hierarchical clustering. The remaining data were filtered for a combination of minimum read count, false discovery rate, and fold-change comparison, as indicated in figure legends. Heat maps were generated using Morpheus (https://software.broadinstitute.org/morpheus/) and upstream regulators were determined using Ingenuity Pathways Analysis (Qiagen). Predicted mouse IFN targets were determined using Interferome (http://interferome.its.monash.edu.au/interferome) with default settings. GSEA analysis[32] was done with 1000 permutations by gene set. The expression dataset was generated by filtering out low abundance genes. Hallmark gene sets were obtained from the Broad MSigDB[33]. Additional gene sets were generated from published datasets[9,19,21]. The gene-filtered expression data are available in Supplementary Table 1 and the GSEA analysis results and input files are available in Supplementary Table 2.

Single-cell RNA-seq. Sca-1^{+} or Sca-1^{-} crypt cells were sorted from one mouse infected with *H. polygyrus* for six days. The two resulting cell suspensions (~70,000 cells each) were submitted as separate samples to be barcoded for single-cell RNA-seq using the Chromium Controller (10X Genomics) and the Single Cell 3′ Library Kit v2 (PN-120236/37/62). Resulting libraries were sequenced on a HiSeq 4000 (Illumina) using HiSeq 4000 PE Cluster Kit (PN PE-410-1001) with HiSeq 4000 SBS Kit (150 cycles, PN FC-410-1002), with one sample being loaded per sequencing lane. 19,834 Sca-1^{-} cells (approximately 15,000 reads per cell) and 7,354 Sca1^{+} cells (approximately 40,000 reads per cell) were successfully barcoded and their transcriptomes were sequenced. Raw sequencing data were processed for initial QC analysis and alignment by our sequencing core (Institute of Human Genetics, UCSF) using Cell Ranger software. Further analysis of differential gene expression and unsupervised hierarchical clustering were performed using the Seurat package (v2.0)[34,35]. Samples were merged after read depth correction and the combined dataset was filtered to exclude cells expressing less than 200 genes. Additionally, genes detected in fewer than 10 cells were removed from the analysis. Variable gene expression was assessed in the filtered data set after correction for mitochondrial gene expression. Linear reduction of the data was performed using principle components analysis, focusing on the first 15 principle components determined to be significant to explain variation in the data set via a large permutation test. Cells were clustered using the function FindClusters with a resolution parameter of 2.0. Graphical representation was achieved using the *t*-SNE algorithm, upon which the independently identified clusters were colour coded. Published lists[19,21] of markers for various intestinal crypt cell types were visualized on the clusters with the DoHeatmap function of the Seurat package. For analysis of cluster contribution by Sca-1^{+} or Sca-1^{-} crypt cells (Fig. 4h), cell numbers were first normalized to the total number sequenced for each population. We considered cluster 19 to be an aberrant cluster owing to low cell number (33 of 26,423 total), enrichment for multiple intestinal lineages, and isolation in *t*-SNE analysis. The mean normalized expression values by cluster are available in Supplementary Table 3 and mean normalized expression values presented in the heat maps are available in Supplementary Table 4. The hypergeometric test for enrichment of the fetal gene program signature within each cluster is available in Supplementary Table 5.

Antibodies. The following antibodies (BioLegend) were used for flow cytometry: CD45 (30-F11), CD326/EpCAM (G8.8), CD44 (IM7), Sca-1 (D7), TCRβ (H57), γδTCR (GL3), NK1.1 (PK136), CD90.2 (53-2.1), CD11b (M1/70) and Gr1 (RB6-8C5). For immunofluorescence staining of sections, the following antibodies were used: GFP (GFP-1020, Aves; ab13790, Abcam), Ki67 (Sp6, Thermo Fischer Scientific), E-cadherin (24E10, Cell Signaling Technology), Sca-1 (e13-161.7, Biolegend), Muc2 (SC-15334, Santa Cruz Biotechnology), Mmp7 (AF2967, R&D Systems). EdU was detected using Click-iT Plus EdU Assay Kit (ThermoFisher).

Organoid culture. Cultures from sorted single cells were established as described[36]. In brief, CD45^{-} EpCAM^{+} CD44^{+} Sca-1^{+} and Sca-1^{-} cells were sorted into PBS containing 10% FCS. Cells were re-suspended in GFR, phenol-free Matrigel (Fisher) supplemented with 500 ng/ml EGF (Sigma-Aldrich), 1 μg/ml Noggin (R&D Sytems), 10 μM Jagged-1 peptide (Anaspec) and 10% R-Spondin1 Conditioned Medium (gift of N. Shroyer). Fifty microlitres of Matrigel containing cells was plated in a 24-well cell culture plate, and left to set at 37 °C for 15 min. Pre-warmed 37 °C ENR Medium (Advanced DMEM/F12, 10 mM HEPES, 1X GlutaMAX, 1% Pen/Strep, 1X N-2 Supplement, 1X B-27 Supplement, 1 mM N-Acetylcysteine, 100 ng/ml Noggin, 50 ng/ml EGF, 5% R-Spondin1 Conditioned Media) with 2.5 μM CHIR99021, 2.5 μM Thiazovinin, and 1 μM Jagged-1 peptide was overlaid. Cells were cultured at 37 °C. After three days, the medium was exchanged for ENR medium without CHIR99021, Thiazovinin, or Jagged-1 peptide. Cultures were passaged after eight days and then every 5–7 days thereafter with growth factor-free Matrigel. Cultures were typically analysed at the end of the first passage by imaging and qPCR. Some qPCR experiments were conducted on established cultures, such as in Fig. 4c–e. For in vitro IFN-γ treatment, wild-type organoid lines were prepared from whole crypts and treated three days after passage by exchanging the standard organoid medium with fresh organoid medium

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

containing 5 ng/ml IFN-γ (485-MI, R&D Systems). Twenty-four hours later, organoids were harvested by centrifugation, aspirating the media and Matrigel, and lysed using RLT buffer (Qiagen).

Quantitative PCR. RNA from 5-mm whole tissue (after QIAshredder), sorted cells, or organoids was extracted using RNeasy Mini or Micro Kits (Qiagen). cDNA was synthesized with High Capacity cDNA Reverse Transcription Kits (Applied Biosystems). qPCR reactions were performed using Power SYBR Green (Invitrogen) on an Applied Biosystems StepOnePlus for whole tissue, or iTaq Universal SYBR Green Supermix (Bio Rad) in 384-well plates on a QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher Scientific) for sorted cells and organoids. Primers used for qPCR are listed in Supplementary Table 6.

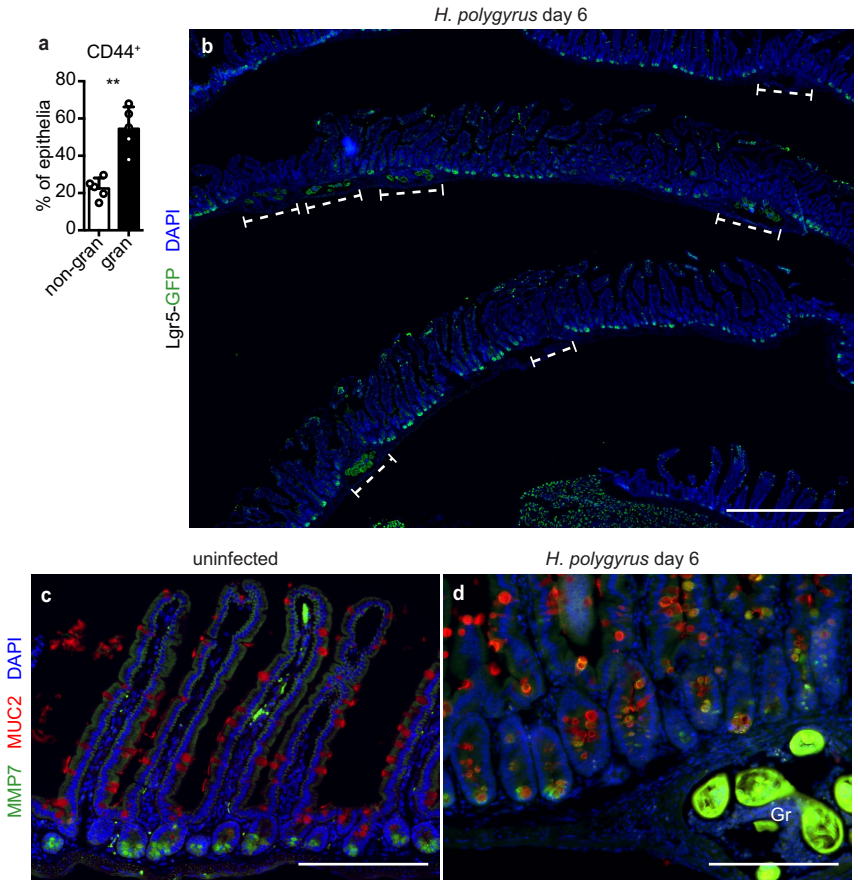
Statistics. Except in Extended Data Fig. 4h, all data points are biological replicates, not formally randomly assigned. The investigators were not formally blinded to allocation during experiments and outcome assessment. All experiments were independently replicated with similar results at least twice, except in Extended Data Fig. 6a, b and the single-cell RNA-seq experiment, which were both performed once. No data were excluded, except in the bulk RNA-seq experiment, as noted in the Methods. No statistical methods were used to predetermine sample size and differences in intra-sample variances were present. Statistical significance was determined in Prism (GraphPad Software) using an unpaired, two-tailed Mann–Whitney test without multiple comparisons correction, except for the use of unpaired, two-tailed *t*-tests in Fig. 3m and Fig. 4c–e, as noted in the legend. Bar charts indicate the mean of samples and error bars represent ± s.d. of the mean. $*P<0.05$, $**P<0.01$, $***P<0.001$, $****P<0.0001$. Hypergeometric tests were performed using GeneProf[37].

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The RNA-seq data reported in this study are available at the Gene Expression Omnibus under accession codes GSE97405 (bulk) and GSE108233 (single-cell). The source data for all charts are available in the online version of the paper.

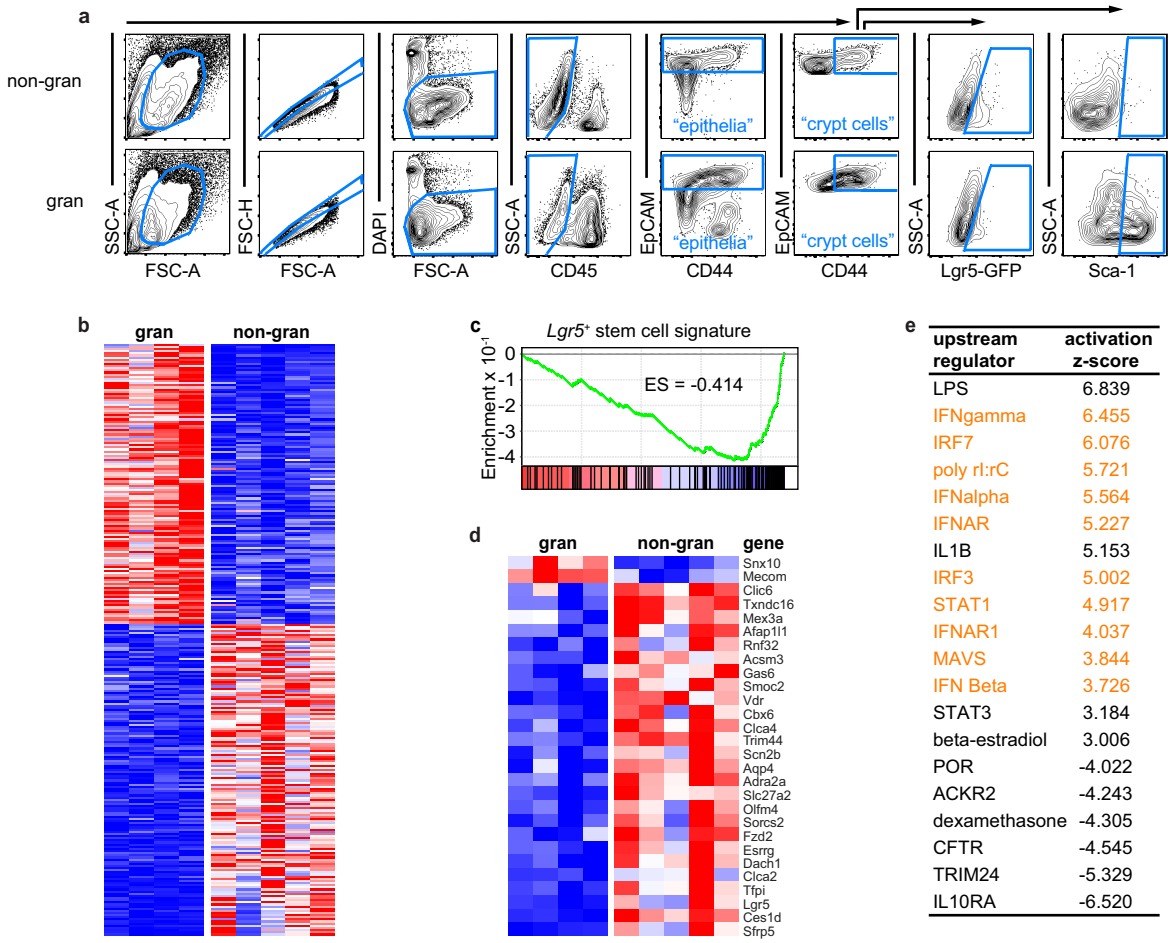
30. Goodyear, A. W., Kumar, A., Dow, S. & Ryan, E. P. Optimization of murine small intestine leukocyte isolation for global immune phenotype analysis. *J. Immunol. Methods* **405**, 97–108 (2014).
31. Gregorieff, A. & Clevers, H. in *Current Protocols in Stem Cell Biology* (John Wiley & Sons, 2015).
32. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
33. Liberzon, A. et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
34. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
35. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
36. Mahe, M. M. et al. Establishment of gastrointestinal epithelial organoids. *Curr. Protoc. Mouse Biol.* **3**, 217–240 (2013).
37. Halbritter, F., Vaidya, H. J. & Tomlinson, S. R. GeneProf: analysis of high-throughput sequencing experiments. *Nat. Methods* **9**, 7–8 (2011).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 1 | Helminth infection alters the crypt and intestinal stem cell niche. Day 6 of *H. polygyrus* infection. **a**, Flow cytometry of $CD44^+$ epithelium from non-granuloma or granuloma biopsies. **b**, Representative image of Lgr5–GFP staining in the duodenum. Granulomas are indicated by the dashed brackets. In some granulomas, the helminth larva is recognizable by autofluorescence. The presence of rare Lgr5–GFP negative crypts is likely to be a sectioning artefact. **c**, **d**, MMP7 and MUC2 staining in normal duodenum or duodenum from mice infected with *H. polygyrus*. Gr, granuloma. $n = 3$ (**b**), or 5 mice (**a**, **c**, **d**). Statistics represent all biological replicates, and all experiments were replicated at least twice. Graphs show mean ± s.d. (**a**). $^{**}P < 0.01$ by unpaired, two-tailed Mann–Whitney test. Scale bars: 1 mm (**b**), 200 μm (**c**, **d**).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

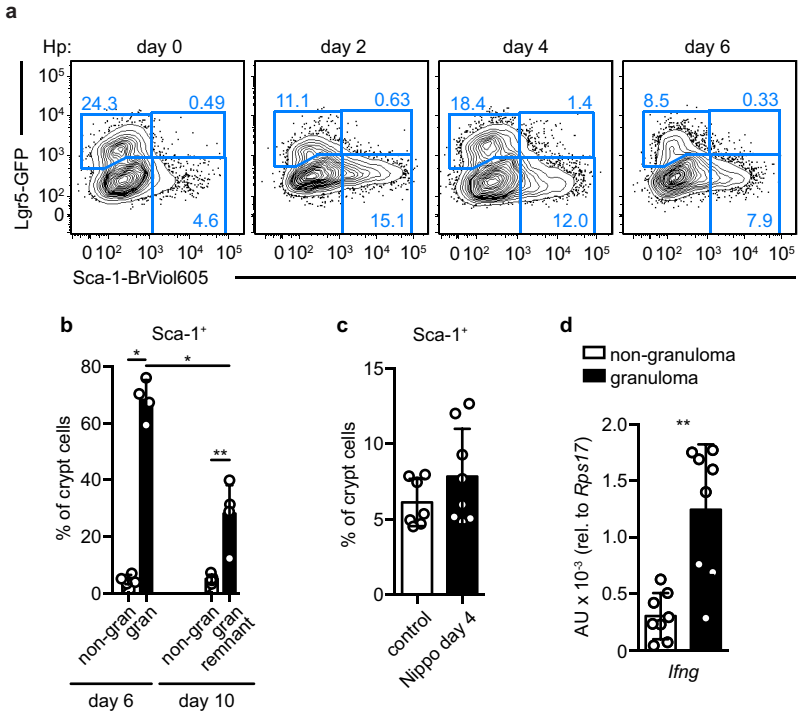


e

upstream regulator	activation z-score
LPS	6.839
IFNgamma	6.455
IRF7	6.076
poly rI:rC	5.721
IFNalpha	5.564
IFNAR	5.227
IL1B	5.153
IRF3	5.002
STAT1	4.917
IFNAR1	4.037
MAVS	3.844
IFN Beta	3.726
STAT3	3.184
beta-estradiol	3.006
POR	-4.022
ACKR2	-4.243
dexamethasone	-4.305
CFTR	-4.545
TRIM24	-5.329
IL10RA	-6.520

Extended Data Fig. 2 | RNA-seq analysis of granuloma-associated crypt epithelium. a, Representative gating example of epithelia, crypt cells, Lgr5–GFP and Sca-1 in biopsied tissue six days after *H. polygyrus* infection. Unfractionated tissue preps (as in Extended Data Fig. 3a) were gated similarly. **b**–**e**, Crypt epithelium was sorted from granuloma and non-granuloma biopsies and subjected to RNA-seq analysis as indicated in the Methods. **b**, The data were filtered for $\geq$100 reads average in either group, FDR $\leq$ 0.05, and fold-change comparison of $\geq$ 2. The 277 genes that passed were compiled into a heat map demonstrating high (red) and low (blue) relative expression. **c**, GSEA for Lgr5$^+$ signature genes[9]. FDR $<$0.01. ES, enrichment score. **d**, Lgr5$^+$ intestinal stem cell signature genes[9] were cross-referenced to the RNA-seq dataset. Data were filtered as in (**b**) except no fold-change requirement was applied. *Clca4* is also known as *Clca3b* **e**, The unfiltered RNA-seq dataset was analysed for upstream regulators using Ingenuity Pathways Analysis. The activation *Z* score indicates the extent of enrichment of targets within the RNA-seq dataset downstream of the indicated regulator, with a positive score indicating enrichment. IFN-related pathways are highlighted in orange. $n = 4$ independently sorted samples (**b**–**e**, granuloma, 20 mice total), or 5 independently sorted samples (**b**–**e**, non-granuloma, 25 mice total).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

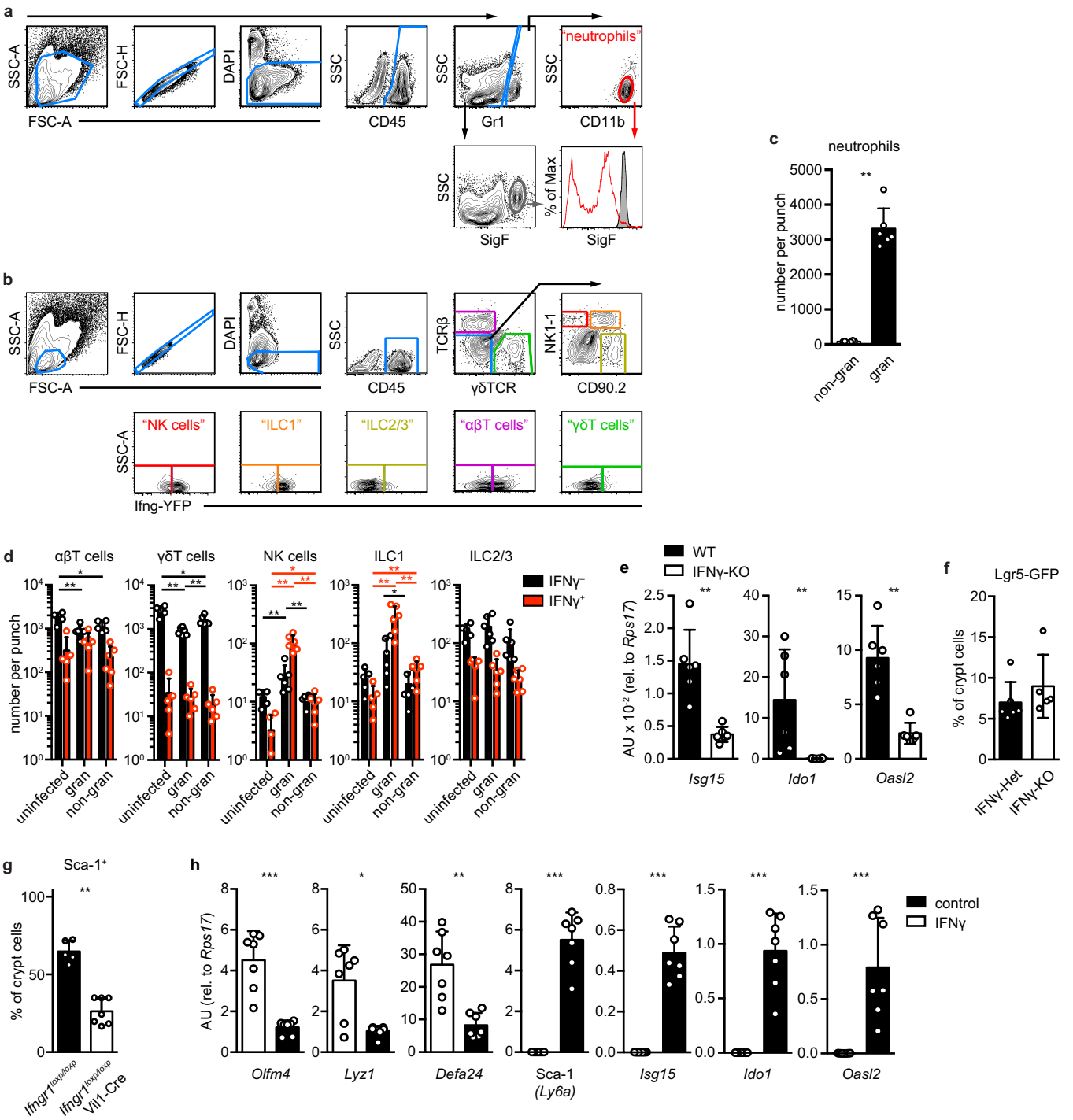


e

gene	fold change (gran/non-gran)	mean granuloma read count	mean non-granuloma read count
Ifna-ps1	NA	0.00	0.00
Ifna1	NA	0.00	0.00
Ifna2	NA	0.00	0.00
Ifna4	NA	0.00	0.00
Ifna5	NA	0.00	0.00
Ifna6	NA	0.00	0.00
Ifna7	NA	0.00	0.00
Ifna9	NA	0.00	0.00
Ifna11	NA	0.00	0.00
Ifna12	NA	0.00	0.00
Ifna13	NA	0.00	0.00
Ifna13	NA	0.00	0.00
Ifna14	NA	0.00	0.00
Ifna15	NA	0.00	0.00
Ifna16	NA	0.00	0.00
Ifnab	NA	0.36	0.00
Ifnb1	NA	0.00	0.00
Ifne	NA	0.00	0.00
Ifng	NA	0.00	0.00
Ifnk	1.18	7.95	6.72
Ifnl2	NA	0.00	0.00
Ifnl3	NA	1.10	0.00
Ifnz	NA	0.56	0.00
Ifnar1	1.11	1412.08	1267.17
Ifnar2	1.22	253.56	207.64
Ifngr1	-1.06	4043.45	4290.54
Ifngr2	-1.12	3368.70	3784.38
Ifnlr1	1.22	513.01	421.71
Il10rb	-1.02	1357.28	1381.12

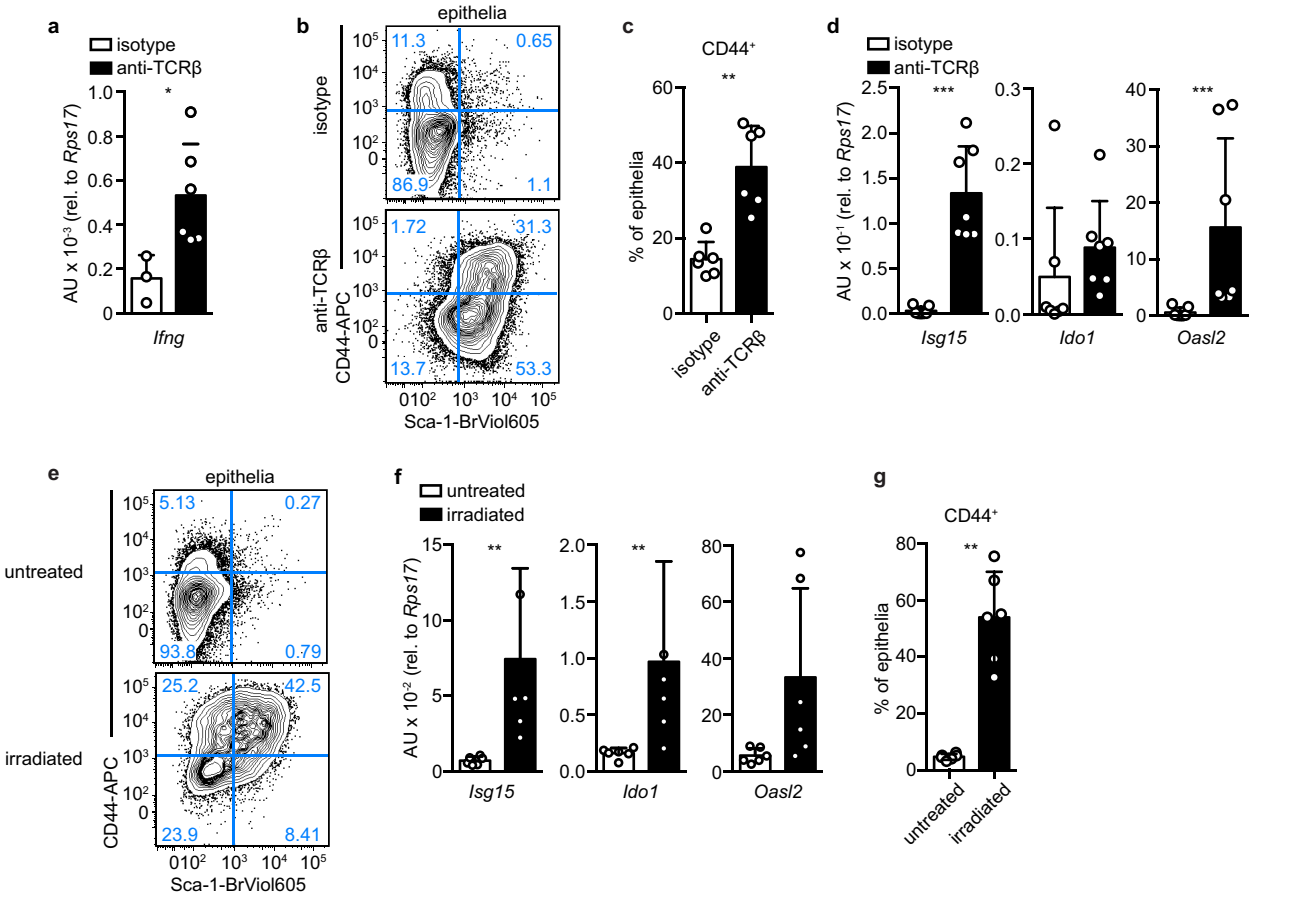
Extended Data Fig. 3 | Sca-1 is expressed in granuloma crypt epithelium and IFN-γ is present in granulomas. a, Representative flow cytometry of Lgr5–GFP and Sca-1 in crypt cells from unfractionated duodenum preps of Lgr5–GFP mice after *H. polygyrus* infection. **b**, Flow cytometry of Sca-1 in crypt cells from biopsies from mice 6 or 10 days after infection with *H. polygyrus*. **c**, Flow cytometry of Sca-1 in crypt cells from unfractionated duodenum preps of mice 4 days after infection with *N. brasiliensis* (Nippo). **d**, Non-granuloma or granuloma biopsies from wild-type mice were analysed by qPCR for *Ifng* transcript. **e**, Fold change and read counts of IFN and IFN receptor genes from RNA-seq performed as in Extended Data Fig. 2b with no filter applied. 'NA' results from division by zero. $n = 4$ (**a**, days 2, 4, 6, **b**, day 6), 5 (**a**, day 0, **b**, day 10), 7 (**c**, controls), or 8 mice (**c**, Nippo, **d**). Statistics represent all biological replicates, and all experiments were replicated at least twice. Graphs show mean ± s.d. (**b**–**d**). $^{*}P < 0.05$, $^{**}P < 0.01$ by unpaired, two-tailed Mann–Whitney test.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



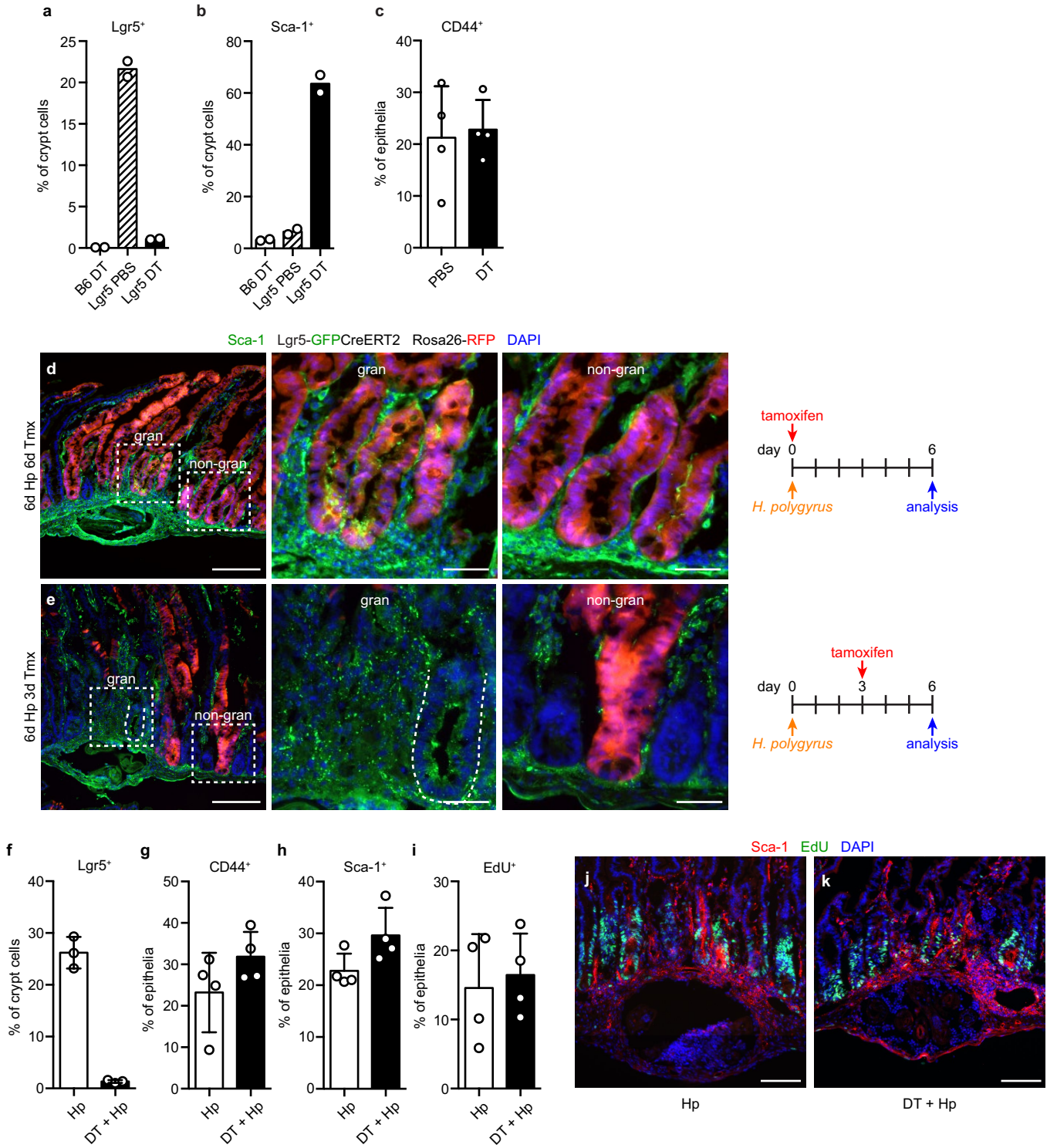
Extended Data Fig. 4 | IFN-γ produced by *H. polygyrus*-responsive immune cells drives the granuloma gene signature. **a**–**g**, Mice were infected with *H. polygyrus* and analysed at day 6, unless otherwise indicated. **a**, **b**, Representative gating example of neutrophils (**a**) and natural killer (NK) cells, ILC1, ILC2/3, αβ T cells, and γδ T cells (**b**). **c**, Neutrophils were enumerated by flow cytometry from non-granuloma (non-gran) or granuloma (gran) biopsies. **d**, *Ifng* reporter mice were untreated (uninfected) or infected (gran or non-gran) with *H. polygyrus* and analysed by flow cytometry 5–6 days later for haematopoietic (CD45$^+$) populations: NK cells, ILC1, ILC2/3, αβ T cells and γδ T cells. No reporter signal was seen in non-lymphoid populations. **e**, Crypt cells were sorted from granuloma biopsies of IFN-γ-knockout (KO) mice and analysed by qPCR for the indicated transcripts. **f**, Lgr5–GFP mice were bred with IFN-γ-knockout (KO) mice and offspring were analysed by flow cytometry for Lgr5–GFP expression in crypt epithelia from granuloma biopsies. **g**, *Ifngr1*$^{loxp/loxp}$ mice were bred with Vil1-Cre mice and analysed by flow cytometry for Sca-1 expression in crypt epithelia from granuloma biopsies. **h**, Wild type organoids were treated with 5 ng ml^{-1} IFN-γ for 24 h and analysed by qPCR for the indicated transcripts. $n = 5$ mice (**d**, uninfected, **f**, KO, **g**, *Ifngr1*$^{loxp/loxp}$), 6 mice (**c**, **d**, infected, **e**, **f**, heterozygous), 7 mice (**g**, *Ifngr1*$^{loxp/loxp}$;Vil1-Cre), or 7 cultures (**h**). Statistics represent all biological replicates, and all experiments were replicated at least twice. Graphs show mean ± s.d. (**c**–**h**). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by unpaired, two-tailed Mann–Whitney test.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



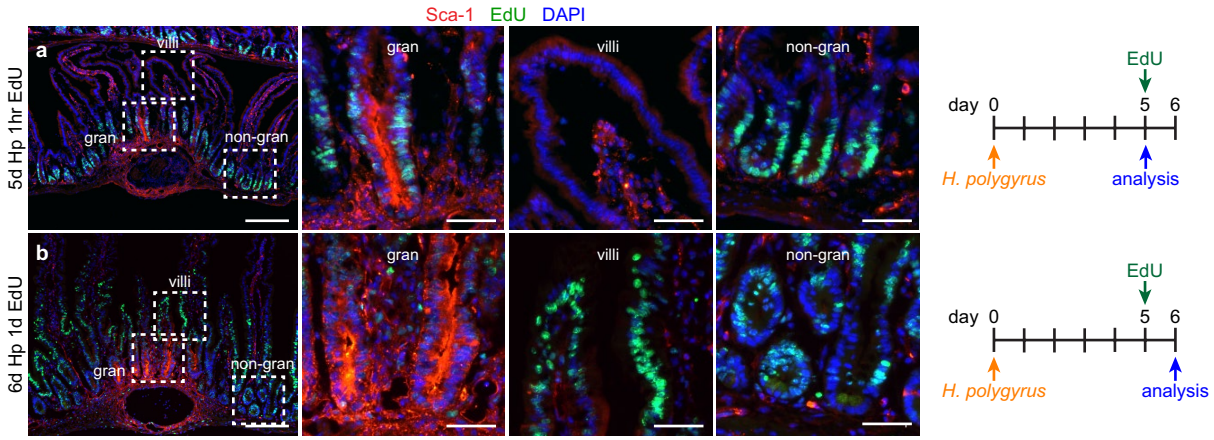
Extended Data Fig. 5 | Inflammation via immune cell activation and irradiation induces granuloma-like epithelial responses. a–d, Mice were treated with 20 μg isotype antibody or anti-TCRβ antibody (clone H57) and analysed 24 h later. **a**, Unfractionated tissue analysed by qPCR for *Ifng* transcript. **b**, Representative flow cytometry of CD44 and Sca-1 in total epithelium. **c**, Epithelium was assessed for crypt size by flow cytometry using frequency of CD44. **d**, Crypt cells were sorted and analysed by qPCR for the indicated transcripts. **e–g**, Mice were untreated or subjected to 10 Gy irradiation and analysed three days later. **e**, Representative flow cytometry of CD44 and Sca-1 expression on total epithelium. **f**, Crypt cells were sorted and analysed by qPCR for the indicated transcripts. **g**, Flow cytometry of the frequency of $CD44^+$ crypt cells among total epithelium. $n = 3$ (**a**, isotype), 6 (**a**, anti-TCRβ, **b**, **c**, **e–g**), or 7 mice (**d**). Statistics represent all biological replicates, and all experiments were replicated at least twice. Graphs show mean ±s.d. (**a**, **c**, **d**, **f**, **g**). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by unpaired, two-tailed Mann–Whitney test.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



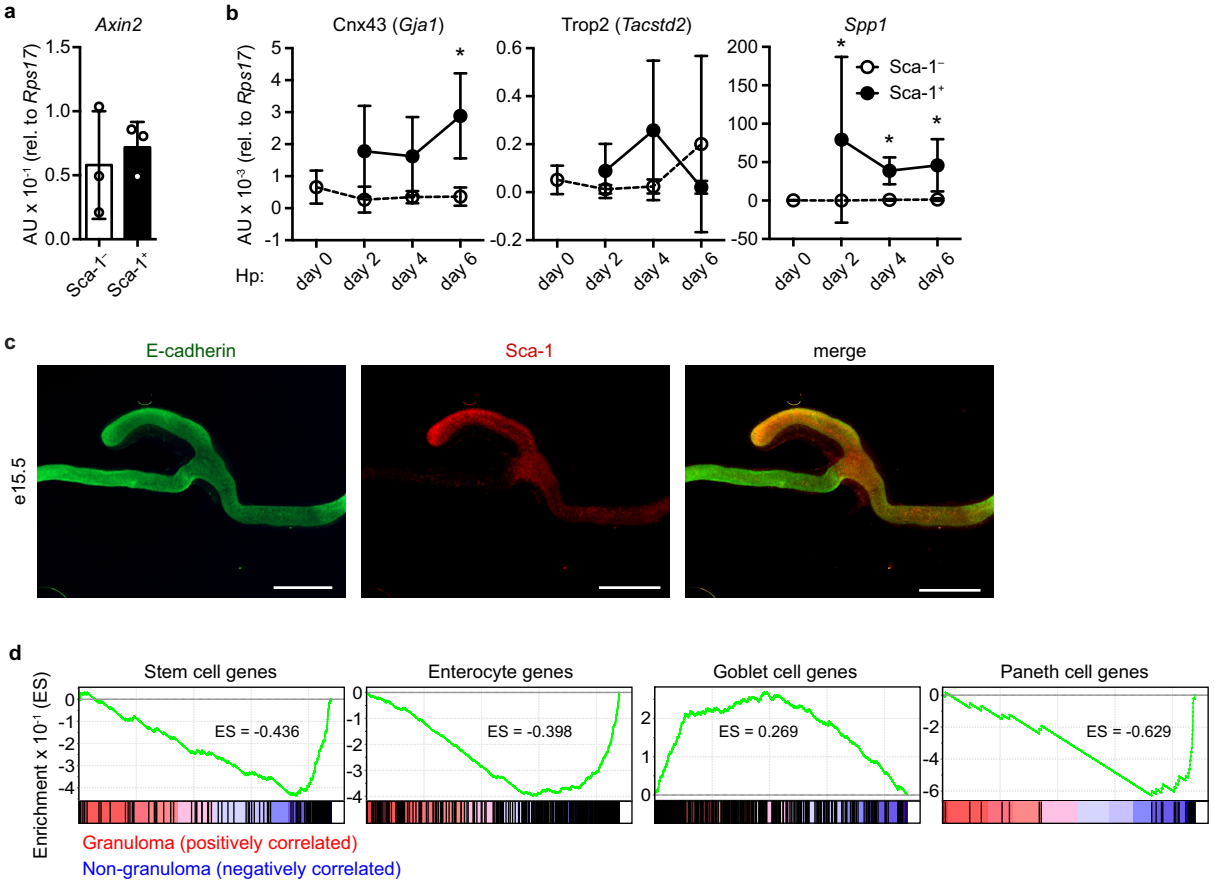
Extended Data Fig. 6 | Granuloma crypt epithelium arises from pre-existing Lgr5$^+$ cells but does not require Lgr5$^+$ cells. a–c, *Lgr5*$^{DTRGFP/+}$ (Lgr5) or wild-type (B6) mice were treated with diphtheria toxin (DT) and analysed a day later by flow cytometry for Lgr5–GFP (**a**), Sca-1 (**b**), or frequency of crypt cells among total epithelium (**c**). **d**, **e**, Representative images of lineage tracing of Lgr5$^+$ precursors and Sca-1 staining in crypts overlying (gran) and adjacent to (non-gran) *H. polygyrus* (Hp) granulomas. *Lgr5*$^{GFP\text{-}creERT2/+}$ *Rosa26*$^{RFP/+}$ mice were administered 2.5 mg tamoxifen (Tmx) either immediately before (**d**) or three days after (**e**) infection with *H. polygyrus*. Mice were analysed at day 6. **f–k**, *Lgr5*$^{DTRGFP/+}$ mice were treated with diphtheria toxin immediately before infection with *H. polygyrus* and analysed by flow cytometry at day 1 for Lgr5–GFP (**f**), or at day 6 for CD44 (**g**), Sca-1 (**h**) and EdU (**i**) in epithelial cells from granuloma biopsies. **j**, **k**, Representative images of Sca-1 and EdU detection at day 6. $n = 2$ (**a**, **b**, **e**), 3 (**d**, **f**, **j**, **k**), or 4 mice (**c**, **g–i**). Experiments were replicated at least twice, except the experiment in **a** and **b**, which was performed once. Graphs show mean ± s.d. (**c**, **f–i**). Scale bars: main, 200 μm; insets 50 μm (**d**, **e**), 100 μm. (**j**, **k**).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



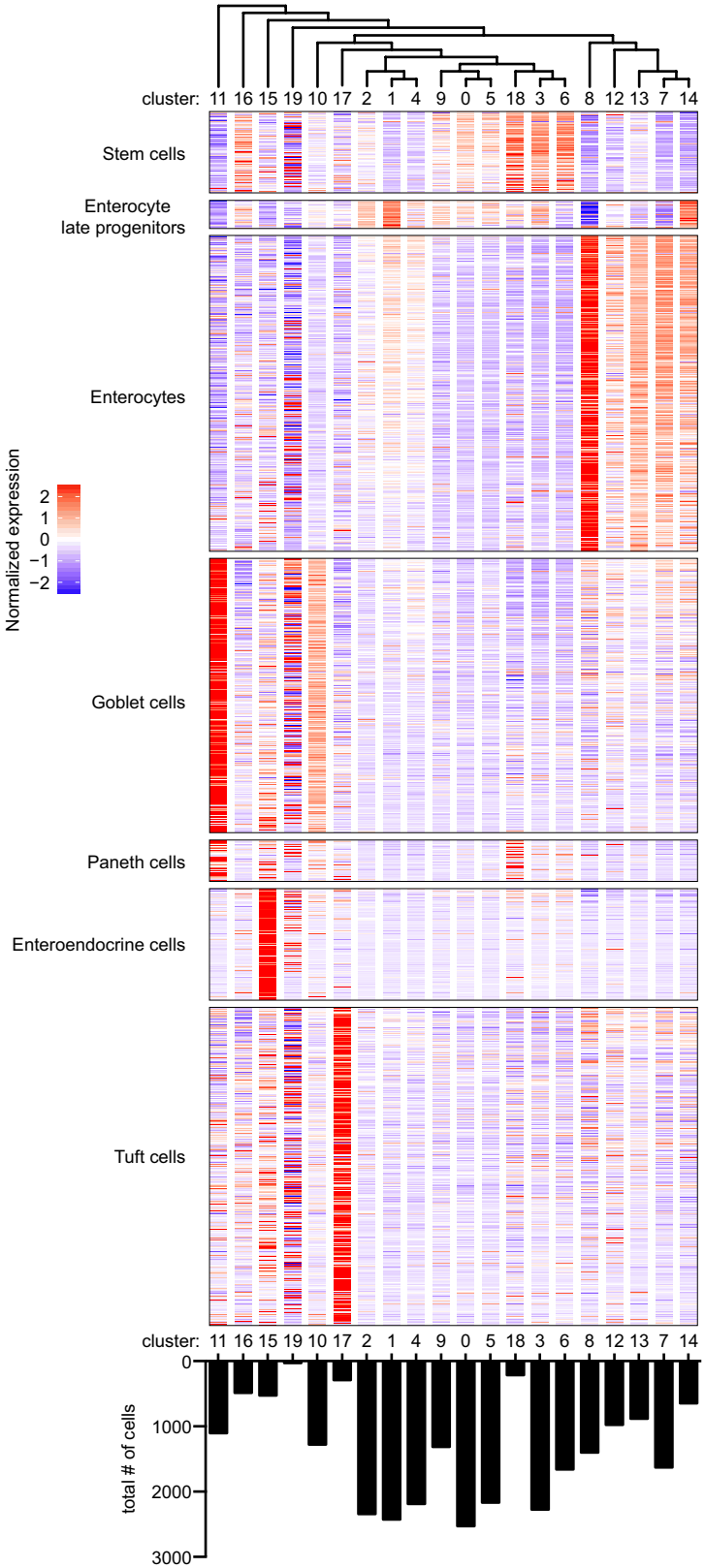
Extended Data Fig. 7 | Granuloma crypt epithelium contributes to epithelial turnover. a, **b**, Representative images of Sca-1 and EdU staining. Wild-type mice were injected with EdU at day 5 of infection and analysed after 1 h (**a**) or 24 h (**b**) to localize labelled cells within villi, and within crypts overlying (gran) or adjacent to (non-gran) *H. polygyrus* granulomas. $n = 4$ (**a**) or 6 mice (**b**). Experiments were replicated at least twice. Scale bars: left, 200 μm; insets, 50 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 8 | Granuloma crypt epithelium activates a fetal-like program and exhibits altered differentiation. a, Material from the cultures described in Fig. 4a, b was analysed by qPCR for *Axin2* transcript. **b**, Sca-1$^+$ or Sca-1$^-$ crypt cells were sorted from mice infected with *H. polygyrus* for the indicated times and analysed by qPCR for fetal transcripts. **c**, Representative image of whole-mount e15.5 fetal intestine stained for Sca-1 and E-cadherin. **d**, Bulk RNA-seq data (as in Fig. 1e) were analysed by GSEA for intestinal epithelial signature genes[21]. Enrichment score (ES) is indicated and all analyses have FDR $<10^{-3}$. $n = 3$ mice (**a**), 3 fetuses (**c**), 4 mice (**b**, day 2, 4, 6), 4 independently sorted samples (**d**, granuloma, 20 mice total), 5 mice (**b**, day 0), or 5 independently sorted samples (**d**, non-granuloma, 25 mice total). Statistics represent all biological replicates, and all experiments were replicated at least twice. Graphs show mean ± s.d. (**a**, **b**). *$P < 0.05$ by unpaired, two-tailed Mann–Whitney test. Scale bar: 1 mm (**c**).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 9 | Markers of adult intestinal cell types in single-cell RNA-seq of Sca-1$^+$ and Sca-1$^-$ crypt epithelium. Clusters identified by unsupervised hierarchical clustering were arranged per the unsupervised dendrogram of cluster relatedness (top) and normalized expression values for intestinal cell type gene signatures[21] were displayed as a heat map in each cluster (middle). The total number of cells in each cluster is shown (bottom).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

https://doi.org/10.1038/s41586-018-0253-5

Induction of innate immune memory via microRNA targeting of chromatin remodelling factors

John J. Seeley[1], Rebecca G. Baker[1], Ghait Mohamed[2], Tony Bruns[2,3], Matthew S. Hayden[1,4], Sachin D. Deshmukh[2], Daniel E. Freedberg[5] & Sankar Ghosh[1]*

Prolonged exposure to microbial products such as lipopolysaccharide can induce a form of innate immune memory that blunts subsequent responses to unrelated pathogens, known as lipopolysaccharide tolerance. Sepsis is a dysregulated systemic immune response to disseminated infection that has a high mortality rate. In some patients, sepsis results in a period of immunosuppression (known as 'immunoparalysis')[1] characterized by reduced inflammatory cytokine output[2], increased secondary infection[3] and an increased risk of organ failure and mortality[4]. Lipopolysaccharide tolerance recapitulates several key features of sepsis-associated immunosuppression[5]. Although various epigenetic changes have previously been observed in tolerized macrophages[6–8], the molecular basis of tolerance, immunoparalysis and other forms of innate immune memory has remained unclear. Here we perform a screen for tolerance-associated microRNAs and identify miR-221 and miR-222 as regulators of the functional reprogramming of macrophages during lipopolysaccharide tolerization. Prolonged stimulation with lipopolysaccharide in mice leads to increased expression of miR-221 and mir-222, both of which regulate brahma-related gene 1 (*Brg1*, also known as *Smarca4*). This increased expression causes the transcriptional silencing of a subset of inflammatory genes that depend on chromatin remodelling mediated by SWI/SNF (switch/sucrose non-fermentable) and STAT (signal transducer and activator of transcription), which in turn promotes tolerance. In patients with sepsis, increased expression of miR-221 and miR-222 correlates with immunoparalysis and increased organ damage. Our results show that specific microRNAs can regulate macrophage tolerization and may serve as biomarkers of immunoparalysis and poor prognosis in patients with sepsis.

Lipopolysaccharide (LPS) tolerance is an immunosuppressive form of innate immune memory that can be modelled in vitro by prolonged treatment of bone-marrow-derived macrophages (BMDMs) with LPS (Extended Data Fig. 1a). As a result of this functional reprogramming of macrophages, a majority of LPS-induced genes are transcriptionally silenced (that is, tolerized) and are not expressed upon re-stimulation[7,9] (Extended Data Fig. 1b). Using this in vitro model (Extended Data Fig. 1c–e) we identified microRNAs (miRNAs) with expression patterns that correlate with tolerance (Fig. 1a). We validated these findings using qPCR (Extended Data Fig. 1f, g) and found that several miRNAs are differentially expressed during tolerance but not during an acute LPS response. Levels of miR-222, in particular, increased late during the LPS response (Extended Data Fig. 1g) and correlated with tolerance induction (Fig. 1b). miR-222 was also upregulated to a lesser extent with prolonged stimulation with tumour necrosis factor (TNF) or interleukin-1β (IL-1β) (Extended Data Fig. 1h), which has been shown to weakly induce innate immune tolerance[10,11]. Pre-treatment of BMDMs with interferon gamma (IFNγ), which inhibits LPS tolerance[8], prevented LPS-induced upregulation of miR-222 (Extended Data Fig. 1i). Although miR-221 is processed from the same primary transcript as miR-222[12], mature levels of miR-221 and of miR-222 do not always correlate (Extended Data Fig. 2a–c). Given that miR-221 is not responsive to LPS (Extended Data Fig. 2a) or IFNγ (Extended Data Fig. 2d) in BMDMs, we focused on miR-222 in BMDM experiments.

BMDMs were transfected with an miR-222 mimic and stimulated with LPS to determine whether miR-222 induced reprogramming independently of other tolerogenic factors (Extended Data Fig. 2e). Overexpression of miR-222 inhibited expression of several inflammatory mediators at the protein (Fig. 1c), mRNA (Extended Data Fig. 2f), and primary transcript level (Extended Data Fig. 2g). Conversely, antagonization of miR-222 resulted in increased inflammatory gene expression, even during a naive LPS response. This effect was relatively mild early after stimulation (data not shown), probably owing to low basal levels of miR-222 expression, but increased in magnitude at later time points (Fig. 1d). To test the effect of miR-222 on tolerance, BMDMs were transduced with an miR-222 antagonist and tolerized in vitro. Antagonization of miR-222 reduced the duration and magnitude of suppression of LPS-response genes (Fig. 1e). In some cases, tolerized cells with antagonized miR-222 produced as much IL-6 or IL-12p40 in response to LPS as did non-tolerized cells (Fig. 1f).

In contrast to other genes, *Tnf* was suppressed at the mRNA level but not at the primary transcript level (Extended Data Fig. 2f, g), which suggests that miR-222 regulates *Tnf* through a mechanism that is distinct from that of other tolerized genes. Indeed, the untranslated region (UTR) of *Tnf* has a predicted binding site for miR-222 (Extended Data Fig. 3a). Luciferase reporter assays (Extended Data Fig. 3b) and CRISPR deletions of the predicted binding site (Extended Data Fig. 3c–g) confirmed that *Tnf* is a target of miR-222. However, the post-transcriptional effects of miR-222 on TNF expression do not contribute to the effects of miR-222 on other genes, as shown by the fact that TNF neutralization did not recapitulate the effects of miR-222 overexpression (Extended Data Fig. 3h, i).

Intact *Tnf* transcription suggested miR-222 does not alter Toll-like receptor 4 (TLR4) signalling. Indeed, miR-222 overexpression did not affect LPS-induced IκBα degradation (Extended Data Fig. 4a–c). We therefore filtered computational predictions for miR-222 targets that were expressed in macrophages, did not affect TLR4 signalling and decreased in expression late in the LPS response (between 8 and 24 h of LPS stimulation; Extended Data Table 1). This approach identified *Brg1* as the target most likely to mediate the transcriptional effects of miR-222 during LPS tolerance. BRG1, a catalytic subunit of the SWI/SNF (BAF) complex, evicts polycomb repressive complexes in an ATP-dependent manner, promoting chromatin accessibility and enabling transcription factor recruitment to specific binding sites[13]. Notably, BRG1 is recruited to the promoters of late LPS-response genes, which require SWI/SNF activity for their transcription[14].

The predicted miR-222–*Brg1* binding site is evolutionarily conserved (Extended Data Fig. 4d), and RNA levels of *Brg1* and miR-222 during the LPS response were inversely correlated (Extended Data

[1]Department of Microbiology & Immunology, College of Physicians & Surgeons, Columbia University, New York, NY, USA. [2]The Integrated Research and Treatment Center for Sepsis Control and Care, Jena University Hospital, Jena, Germany. [3]Department of Internal Medicine IV (Gastroenterology, Hepatology, and Infectious Diseases), Jena University Hospital, Jena, Germany. [4]Section of Dermatology, Department of Surgery, Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA. [5]Department of Medicine, Division of Digestive & Liver Disease, College of Physicians & Surgeons, Columbia University, New York, NY, USA. *e-mail: sg2715@columbia.edu

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

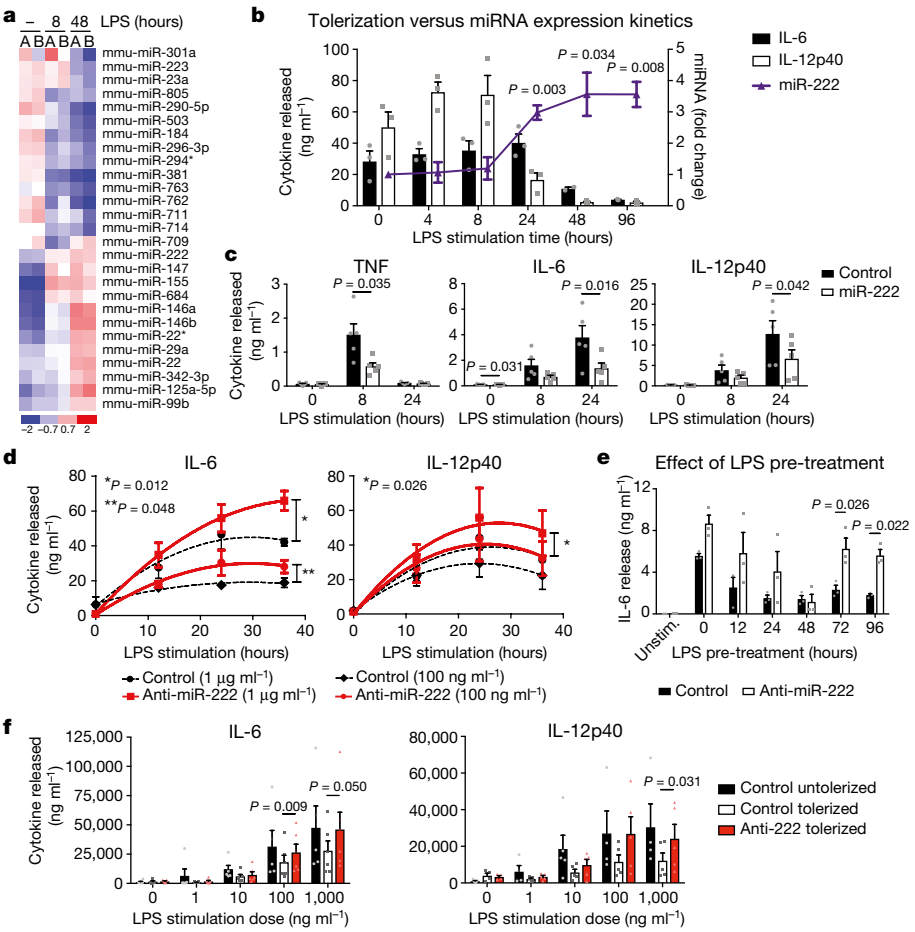


Fig. 1 | miR-222 is upregulated in tolerized BMDMs and suppresses inflammatory gene expression. a, miRNA expression in BMDMs from two mice (labelled A or B) by microarray. **b**, Overlay of qPCR measurement of levels of miR-222 in naive BMDMs (right axis, $n = 4$ biologically independent samples) and cytokine release after re-stimulation of BMDMs as in Extended Data Fig. 1c (left axis, $n = 3$ biologically independent samples) to show the correlation of miR-222 expression kinetics with immunosuppression. **c**, LPS-induced cytokine production after mimic transfection ($n = 5$ biologically independent samples). **d**–**f**, BMDMs (**d**, **f**) or immortalized BMDMs (**e**) were transduced with antagonist constructs. **d**, Cytokine production after stimulation of naive cells ($n = 4$ biologically independent samples). **e**, Re-stimulation of cells with fixed LPS doses after varying pre-treatment time ($n = 3$ independent experiments). **f**, Re-stimulation of cells with varying LPS doses after fixed pre-treatment time ($n = 6$ biologically independent samples). For all graphs, centre value represents mean, and error bars are s.e.m. *P* values calculated by Student's *t*-test (paired, two-sided).

Fig. 4e). Artificial modulation of miR-222 caused an inverse effect on *Brg1* mRNA and protein levels (Extended Data Fig. 4f–h). To confirm that this was due to direct targeting, we cloned the *Brg1* UTR into a luciferase reporter. miR-222 suppressed luciferase activity resulting from co-transfection in a dose-dependent manner only if the miR-222-binding site in the *Brg1* UTR was intact (Extended Data Fig. 4i). We compared the effects of miR-222 overexpression on genes previously identified as being SWI/SNF-dependent in macrophages[15]. Overexpression of miR-222 preferentially suppressed expression of SWI/SNF-dependent genes (Fig. 2a and Extended Data Fig. 4j). Furthermore, BRG1 recruitment to inflammatory gene promoters was reduced after miR-222 overexpression (Fig. 2b). Histone H3 acetylation, which occurs downstream[14] of BRG1 activity, was also reduced (Extended Data Fig. 4k). By contrast, histone H4 acetylation at these promoters, which occurs before BRG1 recruitment[16,17], was unaffected (Extended Data Fig. 4l). Finally, CRISPR–Cas9 disruption of the miR-222-binding site in the *Brg1* UTR in RAW cells (Extended Data Fig. 4m) prevented miR-222-mediated suppression of some SWI/SNF-dependent genes (Extended Data Fig. 4n).

To characterize the biological role of miR-222, we generated a mouse knockout model. However, miR-221 and miR-222 are encoded in the same transcript, are induced by LPS in certain cell types (Extended Data Fig. 2b, c), have similar seed sequences (Extended Data Fig. 5a), have substantial overlap in predicted mRNA targets (Extended Data Fig. 5b) and are both predicted to bind to the same target site in the *Brg1* UTR (Extended Data Fig. 5c). Furthermore, as with miR-222, overexpression of miR-221 downregulates levels of *Brg1* (Extended Data Fig. 5d) and has downstream effects on inflammatory gene expression (Extended Data Fig. 5e). Therefore, we targeted both miRNAs for deletion[18] (Extended Data Fig. 5f–h). We then used qPCR and RNA sequencing to characterize the LPS response in *mir-221 mir-222* knockout macrophages (Fig. 2c). Although the increase in *Brg1* expression in peritoneal macrophages from knockout mice was modest compared to in vitro experiments, *mir-221 mir-222* knockout cells expressed higher levels of many *Brg1*-dependent genes, as well as *Tnf* (Extended Data Fig. 5i, j). Some *Brg1*-dependent genes were more affected by *mir-221 mir-222* knockout than others (compare *Il6* and *Nos2* in Extended Data Fig. 5j), which suggests differential sensitivity to changes in levels of BRG1.

To better understand the mechanisms of altered gene expression in cells that lack miR-221 and miR-222 (Extended Data Fig. 5k), we analysed the promoters of affected genes to identify common regulatory features. Although we obtained similar results in multiple analyses of subsets of affected genes (Extended Data Fig. 6a–f), we limited our main analysis to LPS genes that are most suppressed in tolerized wild-type cells (358 genes out of 1,036 genes in total that are responsive to LPS; Fig. 2d). Roughly half of these 358 genes were expressed at higher levels in tolerized knockout cells compared to tolerized wild-type cells ('de-repressed' genes, Fig. 2e), and roughly half were unaffected

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

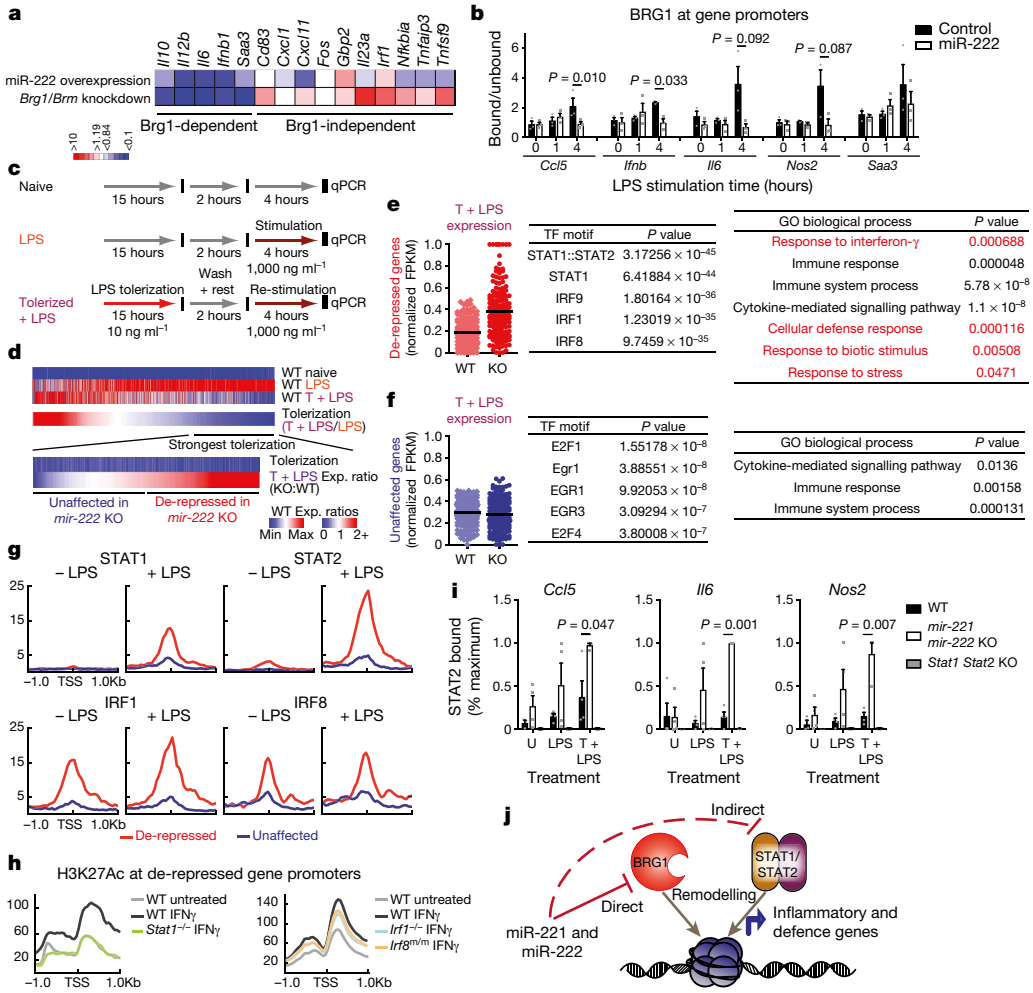


Fig. 2 | miR-222 suppresses BRG1- and STAT-dependent inflammatory gene expression. a, Comparison of miR-222 mimic transfection and the effect of *Brg1* and *Brm* (also known as *Smarca2*) knockdown[15] on LPS-induced gene expression. **b**, ChIP in immortalized BMDMs transduced with overexpression constructs ($n = 3$ independent experiments; *P* values from Students *t*-test for paired values, two-sided). **c**, **d**, Schematic of treatments (**c**) and genes (**d**) analysed in **e**–**i**. Exp., expression; KO, knockout; T, tolerized. **e**, **f**, Dot plot of RNA sequencing expression values (normalized to maximal expression per gene) for wild-type (WT) and *mir-221 mir-222* knockout cells (KO), top five predicted[34] transcription factor motifs, and statistically over-represented Gene Ontology (GO) terms (determined by PANTHER) for indicated gene groups ($n = 103$ gene expression values per group). Gene Ontology terms unique to de-repressed genes are highlighted in red. FKPM, fragments per kilobase per million mapped reads; TF, transcription factor. **g**, **h**, Transcription factor occupancy (**g**) and histone modification (**h**) at promoters, quantified from published ChIP–seq datasets[20–25]. *Irf8*$^{m/m}$ is a homozygous mutant with the hypomorphic *Irf8*^{R294C} allele. TSS, transcription start site. **i**, ChIP for STAT2 occupancy in peritoneal macrophages. Values normalized to maximal binding detected for each ChIP (wild type, *mir-221 mir-222* knockout $n = 4$ biologically independent samples; *Stat1 Stat2* knockout, $n = 2$ biologically independent samples. *P* values were calculated only for wild-type versus *mir-221 mir-222* knockout comparisons, by Student's *t*-test, two-sided, heteroscedastic). **j**, Model of effect of miR-221 and miR-222 on chromatin at affected gene promoters. For all bar graphs and dot plots, centre represents mean and error bars (if present) represent s.e.m.

('unaffected' genes, Fig. 2f). The promoters of de-repressed genes were enriched for the binding motifs of interferon regulatory factors (IRFs) as well as STAT1 and STAT2 (Fig. 2e), whereas those of unaffected genes were enriched for E2F and EGR family motifs (Fig. 2f). An analysis of predicted downstream functions of the de-repressed gene subset found an enrichment for IFN-response genes (Fig. 2e), and LPS-induced expression of many of these genes is reduced in *Ifnar1* knockout cells[19]. This implies that many of these genes are a part of the late LPS response, transcribed as a result of STAT activation by autocrine and/or paracrine signalling by IFN generated from the initial LPS stimulation.

To determine whether the predicted binding motifs were used during the LPS response, we analysed transcription factor occupancy using previously published chromatin immunoprecipitation followed by sequencing (ChIP–seq) data[20–23]. IRF1 and IRF8 were found to be selectively pre-associated with promoters of de-repressed genes (Fig. 2g and Extended Data Fig. 6g). However, STAT1 and STAT2 were recruited specifically to the promoters of de-repressed genes only after LPS stimulation (Fig. 2g). Other transcription factors, such as NF-κB, were not differentially recruited (Extended Data Fig. 6h). Furthermore, in cells with a deletion or mutation of *Irf1* or *Irf8*, respectively[24], cytokine-induced H3K27 (histone H3, lysine 27) acetylation—a marker of active transcription—was diminished at the promoters of de-repressed genes, whereas deletion of *Stat1*[25] almost completely abolished cytokine-induced H3K27 acetylation at these genes (Fig. 2h). Consistent with this analysis, STAT2 recruitment was significantly higher at the promoters of de-repressed genes in tolerized *mir-221 mir-222* knockout cells after re-stimulation, compared to wild-type cells (Fig. 2i). Furthermore, levels of *Stat1* mRNA are higher in *mir-221 mir-222* knockout cells and in cells in which *Brg1* is overexpressed (Extended Data Fig. 7i, j) than they are in wild-type cells. Therefore, miR-221 and miR-222 perturb SWI/SNF promoter recruitment, which leads to the repression of STAT activity at inflammatory gene promoters. As BRG1 and STAT transcription factors work cooperatively only at certain gene promoters to enable IFN- and cytokine-induced gene transcription[26,27], miR-221 and miR-222 may limit expression of specific genes (Fig. 2i).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

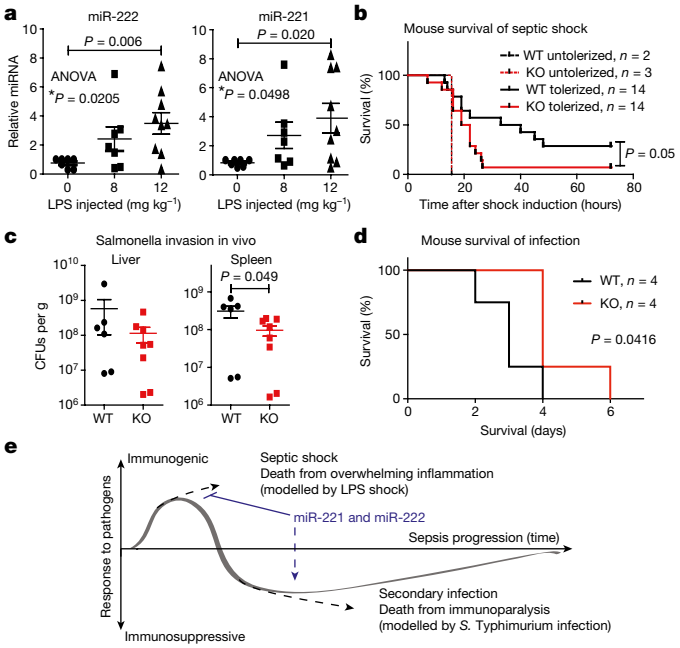


Fig. 3 | miR-221 and miR-222 protect against inflammatory septic shock but increase susceptibility to live infection in mice. a, miRNA levels in blood buffy coat 24 h after LPS injection ($n = 7$ mice for dose 0, 7 mice for dose 8 mg per kg and 9 mice for dose 12 mg per kg). **b**, Survival of littermates that were untreated or tolerized before lethal LPS injection, as in Extended Data Fig.. **c**, **d**, Bacterial dissemination (**c**, $n = 6$ wild-type and 8 knockout mice) and host survival (**d**) after intraperitoneal injection with *S.* Typhimurium. **e**, Model of miR-221 and miR-222 effect on the immune response and host survival during the course of sepsis. For **a**, **c**, *P* values were determined by Student's *t*-test (paired, two-sided). *P* values for **b**, **d**, log-rank (Mantel–Cox) test (performed only for $n > 3$). For all dot plots, centre line represents mean; error bars represent s.e.m.

Next, we examined miR-221 and miR-222 activity using a model of sterile inflammatory shock induced by a high-dose LPS injection. In this system, changes that decrease inflammation increase survival: therefore, we used this model mainly to determine whether the anti-inflammatory effects of miR-221 and miR-222 we observe in vitro also occur in vivo. After LPS injection, levels of miR-221 and miR-222 in circulating immune cells were elevated (Fig. 3a). To determine whether this is physiologically relevant, LPS tolerance was induced in wild-type and *mir-221 mir-222* knockout littermates by administering two sublethal doses of LPS before a lethal LPS dose: this regimen induces sufficient tolerance to prevent lethality in wild-type mice (Extended Data Fig. 7a, b). Although *mir-221 mir-222* knockout mice were also protected from lethality, the *mir-221 mir-222* knockout mice exhibited more symptoms of septic shock (Extended Data Fig. 7c), which indicates decreased anti-inflammatory effects in the knockouts. To test whether miR-221 and miR-222 contribute to survival under more extreme conditions, we used a model of septic shock in which tolerance is only partially protective against lethality (Extended Data Fig. 7d, e). In this model, absence of miR-221 and miR-222 decreased the median survival time (from 36.5 h to 20.5 h) as well as the likelihood of septic shock survival over a 72-h period (Fig. 3b).

Although LPS-induced septic shock is used to study acute inflammation in vivo, this model does not recapitulate sepsis in patients or necessarily predict the effect of inflammatory regulators on patient outcome. Therefore, to study the role of miR-221 and miR-222 in a model that better reflects the systemic innate response to pathogen challenge, we used a *Salmonella enterica* subsp. *enterica* serovar Typhimurium (*S.* Typhimurium) infection model. First, we performed in vitro assays using green fluorescent protein (GFP)-expressing *S.* Typhimurium to infect BMDMs. BMDMs from *mir-221 mir-222* knockout mice exhibited increased GFP per cell soon after infection (Extended Data Fig. 7f–h). At later time points, this difference was not observed (Extended Data Fig. 7h), which suggests that—despite increased phagocytosis—*mir-221 mir-222* knockout cells are more efficient at suppressing intracellular replication and/or survival. We confirmed this finding by lysing BMDMs and comparing bacterial colony-forming unit (CFU) recovery at early and late time points after infection (Extended Data Fig. 7i). To test miR-221 and miR-222 effects in vivo, wild-type and knockout mice were injected intraperitoneally with the same strain of *S.* Typhimurium. Two days after infection, fewer bacterial CFUs were recovered from the liver and spleen of *mir-221 mir-222* knockout mice (Fig. 3c). In addition, *mir-221 mir-222* knockout mice exhibited increased survival time (Fig. 3d), suggesting that the loss of miR-221 and miR-222 confers resistance to bacterial replication and/or dissemination. These findings suggest that miR-221 and miR-222 broadly suppress inflammation and innate immune function. During the early stages of sepsis expression of miR-221 and miR-222 may be protective, by limiting the excessive inflammatory cytokine production that contributes to septic shock. Conversely, miR-221 and miR-222 appear to contribute to immunoparalysis, and increased miR-221 and miR-222 expression may enhance lethality during the later stages of sepsis (Fig. 3e).

Because it is unclear which models most accurately resemble conditions in patients, we next examined miR-221 and miR-222 expression in human disease. Consistent with results from mouse cells, miR-221 and miR-222 are both upregulated in response to prolonged LPS stimulation of a human monocyte-like cell line, whereas only miR-222 is upregulated by LPS in this cell line after phorbol 12-myristate 13-acetate (PMA)-induced differentiation to a macrophage-like cell type (Extended Data Fig. 8a, b). Next we analysed miR-221 and miR-222 expression in three patient cohorts. In the first cohort (Extended Data Fig. 8c), we quantified miR-221 and miR-222 levels in peripheral blood mononuclear cells from ten sequential patients from an intensive care unit, who met sepsis criteria[28] within 4 h of admission to the intensive care unit. Compared to peripheral blood mononuclear cells from healthy donors, miR-221 and miR-222—but not several other inflammation-associated miRNAs—were significantly higher in samples from patients in the intensive care unit (Fig. 4a). Levels of miR-221 and miR-222 expression were then examined in a second cohort of patients, all of whom had acute decompensated liver disease and clinical suspicion of infection (Extended Data Fig. 8d). Patients with organ failure, as defined by the chronic liver failure–sequential organ failure assessment, had significantly higher levels of miR-222 than patients without such failure (Fig. 4b). Levels of miR-221 correlated with those of miR-222 (Extended Data Fig. 8f), but were not increased to statistically significant levels (Fig. 4c). Levels of miR-222 in this cohort inversely correlated with levels of *BRG1* expression (Fig. 4d). In a set of matched peripheral blood mononuclear cell and serum samples, levels of miR-222 and TNF were also inversely correlated (Fig. 4e). Finally, the inverse correlation between miR-222 and *BRG1* was also observed in $CD14^+$ monocytes sorted from the peripheral blood mononuclear cell population of a third clinical cohort (Fig. 4f and Extended Data Fig. 8e), confirming changes in levels of miR-222 and *BRG1* in myeloid cells.

Unlike generalized inflammatory markers, miR-222 elevation correlates specifically with severe sepsis. miR-222 levels do not correlate with inflammatory markers such as C-reactive protein or white blood cell count, but showed a significant correlation with organ-damage markers, including creatinine and the model for end-stage liver disease score (Extended Data Fig. 8g–j). miR-222 expression may be a useful biomarker for detecting patients who are undergoing septicaemia-induced immunoparalysis and are, therefore, predisposed to organ failure and mortality.

In summary, our data establish a model in which miR-221 and miR-222 restrict chromatin remodelling and silence transcription to enforce innate immune tolerance. After prolonged innate immune signalling, increased expression of miR-221 and miR-222 reduces expression of BRG1, which leads to changes in SWI/SNF complex levels or composition that result in the selective expression of only those LPS-response genes with the most favourable chromatin states. The fact that substantial changes in gene expression result from modest miR-221- and miR-222-dependent

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

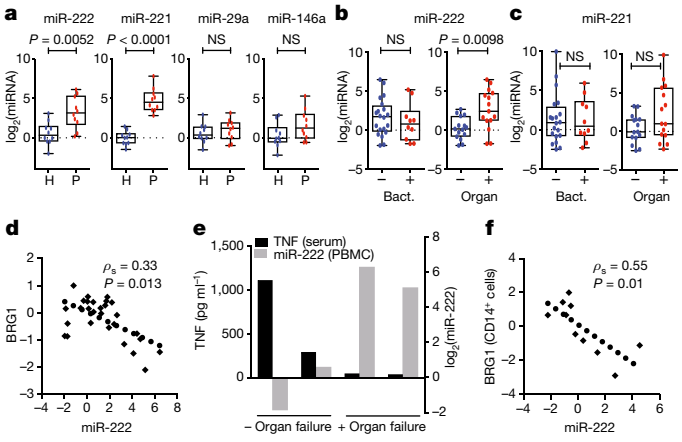


Fig. 4 | miR-222 correlates with immunosuppression and severe sepsis in patients. a, miRNA expression in peripheral blood mononuclear cells from healthy donors (H) or patients (P) from the intensive care unit with sepsis ($n = 10$ per group). *P* values, two-sided Mann–Whitney *U* test. **b**, **c**, Expression in peripheral blood mononuclear cells from a cohort of patients with chronic liver disease, stratified for bacterial infection ($n = 20$ patients negative (−), 10 positive (+)) and inflammation-related organ failure (acute-on-chronic liver failure; $n = 15$ patients per group). *P* values, two-sided Mann–Whitney *U* test. Bact., cells from patients with bacterial infection; organ, cells from patients with organ failure. **d**, **f**, Correlation of RNA levels in peripheral blood mononuclear cells (**d**, $n = 28$ patients) or $CD14^+$ monocytes of patients with multiple organ failure (**f**, $n = 10$ patients). Spearman's rho (ρ) and *P* values from bivariate non-parametric regression analysis. **e**, Correlation of serum TNF and miR-222 expression in peripheral blood mononuclear cells (PBMC) from four patients with signs of infection. Two patients had organ failure according to the EASL CLIF-C criteria for acute-on-chronic liver failure. In box and whisker plots, median is shown as centre line; box, 25th to 75th percentiles; whiskers, minimum to maximum values. NS, not significant.

changes in BRG1 expression is consistent with previous reports that the mutation or deletion of a single allele of the SWI/SNF subunit is sufficient to confer strong phenotypic effects[29,30]. Hence, by fine-tuning the levels of BRG1, miR-221 and miR-222 can prevent prolonged expression of STAT-dependent inflammatory genes in macrophages, thereby leading to tolerance or innate immunoparalysis (Extended Data Fig. 9). By contrast, robust activation of STAT1—for example, by co-stimulation with IFNγ—can block[8] or even reverse[31,32] LPS tolerance and innate immunoparalysis. Consistent with such a role for STAT1, treatment with IFNγ has been shown to improve outcomes in sepsis[33].

Although LPS tolerance promotes survival in mouse models of sterile shock, patients with sepsis probably succumb to primary or secondary[1] infections owing to immunosuppression as a result of functional reprogramming of myeloid cells. Thus, the same innate immunoparalysis that is protective in the mouse LPS-shock model would be responsible for organ damage and mortality in human patients with sepsis. We identify miR-221 and miR-222 as mediators of tolerance and show that miR-221 and miR-222 expression may distinguish patients with organ failure who are at high risk of mortality from those patients with infection alone. Thus, the monitoring of miR-221 and miR-222, or related bio-markers, may help clinicians to stratify patients with sepsis into groups on the basis of whether they would benefit from pro-inflammatory immunotherapies or classical anti-inflammatory treatments.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0253-5.

Received: 28 February 2017; Accepted: 3 May 2018;
Published online: 27 June 2018

1. Hotchkiss, R. S., Monneret, G. & Payen, D. Sepsis-induced immunosuppression: from cellular dysfunctions to immunotherapy. *Nat. Rev. Immunol.* **13**, 862–874 (2013).
2. Ertel, W. et al. Downregulation of proinflammatory cytokine release in whole blood from septic patients. *Blood* **85**, 1341–1347 (1995).
3. Otto, G. P. et al. The late phase of sepsis is characterized by an increased microbiological burden and death rate. *Crit. Care* **15**, R183 (2011).
4. Boomer, J. S. et al. Immunosuppression in patients who die of sepsis and multiple organ failure. *J. Am. Med. Assoc.* **306**, 2594–2605 (2011).
5. Cavaillon, J. M. & Adib-Conquy, M. Bench-to-bedside review: endotoxin tolerance as a model of leukocyte reprogramming in sepsis. *Crit. Care* **10**, 233 (2006).
6. Saeed, S. et al. Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science* **345**, 1251086 (2014).
7. Foster, S. L., Hargreaves, D. C. & Medzhitov, R. Gene-specific control of inflammation by TLR-induced chromatin modifications. *Nature* **447**, 972–978 (2007).
8. Chen, J. & Ivashkiv, L. B. IFN-γ abrogates endotoxin tolerance by facilitating Toll-like receptor-induced chromatin remodeling. *Proc. Natl Acad. Sci. USA* **107**, 19438–19443 (2010).
9. Mages, J., Dietrich, H. & Lang, R. A genome-wide analysis of LPS tolerance in macrophages. *Immunobiology* **212**, 723–737 (2008).
10. Fraker, D. L., Stovroff, M. C., Merino, M. J. & Norton, J. A. Tolerance to tumor necrosis factor in rats and the relationship to endotoxin tolerance and toxicity. *J. Exp. Med.* **168**, 95–105 (1988).
11. Cavaillon, J.-M., Pitton, C. & Fitting, C. Endotoxin tolerance is not a LPS-specific phenomenon: partial mimicry with IL-1, IL-10 and TGFβ. *J. Endotoxin Res.* **1**, 21–29 (1994).
12. Di Leva, G. et al. MicroRNA cluster 221–222 and estrogen receptor α interactions in breast cancer. *J. Natl. Cancer Inst.* **102**, 706–721 (2010).
13. Stanton, B. Z. et al. *Smarca4* ATPase mutations disrupt direct eviction of PRC1 from chromatin. *Nat. Genet.* **49**, 282–288 (2017).
14. Ramirez-Carrozzi, V. R. et al. Selective and antagonistic functions of SWI/SNF and Mi-2β nucleosome remodeling complexes during an inflammatory response. *Genes Dev.* **20**, 282–296 (2006).
15. Ramirez-Carrozzi, V. R. et al. A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138**, 114–128 (2009).
16. Saccani, S., Pantano, S. & Natoli, G. Two waves of nuclear factor κB recruitment to target promoters. *J. Exp. Med.* **193**, 1351–1359 (2001).
17. Agalioti, T., Chen, G. & Thanos, D. Deciphering the transcriptional histone acetylation code for a human gene. *Cell* **111**, 381–392 (2002).
18. Park, C. Y. *et al.* A resource for the conditional ablation of microRNAs in the mouse. *Cell Reports* **1**, 385–391 (2012).
19. Ourthiague, D. R. et al. Limited specificity of IRF3 and ISGF3 in the transcriptional innate-immune response to double-stranded RNA. *J. Leukoc. Biol.* **98**, 119–128 (2015).
20. Mancino, A. et al. A dual *cis*-regulatory code links IRF8 to constitutive and inducible gene expression in macrophages. *Genes Dev.* **29**, 394–408 (2015).
21. Tong, A. J. et al. A stringent systems approach uncovers gene-specific mechanisms regulating inflammation. *Cell* **165**, 165–179 (2016).
22. Garber, M. et al. A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol. Cell* **47**, 810–822 (2012).
23. Cohen, M. et al. Chronic exposure to TGFβ1 regulates myeloid cell inflammatory response in an IRF7-dependent manner. *EMBO J.* **33**, 2906–2921 (2014).
24. Langlais, D., Barreiro, L. B. & Gros, P. The macrophage IRF8/IRF1 regulome is required for protection against infections and is associated with chronic inflammation. *J. Exp. Med.* **213**, 585–603 (2016).
25. Ostuni, R. et al. Latent enhancers activated by stimulation in differentiated cells. *Cell* **152**, 157–171 (2013).
26. Huang, M. et al. Chromatin-remodelling factor BRG1 selectively activates a subset of interferon-α-inducible genes. *Nat. Cell Biol.* **4**, 774–781 (2002).
27. Ni, Z. et al. Apical role for BRG1 in cytokine-induced promoter assembly. *Proc. Natl Acad. Sci. USA* **102**, 14611–14616 (2005).
28. Levy, M. M. et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Intensive Care Med.* **29**, 530–538 (2003).
29. Kadoch, C. et al. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45**, 592–601 (2013).
30. Bultman, S. et al. A *Brg1* null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol. Cell* **6**, 1287–1295 (2000).
31. Cheng, S. C. et al. Broad defects in the energy metabolism of leukocytes underlie immunoparalysis in sepsis. *Nat. Immunol.* **17**, 406–413 (2016).
32. Leentjens, J. et al. Reversal of immunoparalysis in humans *in vivo*: a double-blind, placebo-controlled, randomized pilot study. *Am. J. Respir. Crit. Care Med.* **186**, 838–845 (2012).
33. Döcke, W. D. et al. Monocyte deactivation in septic patients: restoration by IFN-γ treatment. *Nat. Med.* **3**, 678–681 (1997).
34. Zambelli, F., Pesole, G. & Pavesi, G. Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res.* **37**, W247–W252 (2009).

Acknowledgements We thank M. McManus (UCSF) for generously providing us with a targeting construct used in the generation of *mir-221 mir-222* knockout

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

mice. This work was supported by grants R21-AI116082 and R37-AI33443 to S.G. from the National Institutes of Health.

Reviewer information *Nature* thanks G. Crabtree, R.Hotchkiss, M. Netea and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions J.J.S. performed the majority of the experiments and writing of the manuscript. R.G.B. performed the microRNA microarray experiment. G.M. performed experiments on patient samples. T.B., S.D.D. and D.E.F. assisted with experimental design and collected patient samples for the human portions of this study. M.S.H. assisted with experimental design and the writing of the manuscript. S.G. conceived the study, and provided guidance with experimental design and the writing of the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0253-5.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0253-5.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to S.G.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Cell culture. RAW 264.7 cells (ATCC TIB-7) were cultured in DMEM supplemented with 10% fetal bovine serum. 293FT cells (Invitrogen R7007) and L-929 cells (ATCC CCL-1) were cultured in DMEM supplemented with 10% fetal bovine serum. Cells were purchased from vendor and tested for mycoplasma contamination before use (no further authentication of line identity was performed). L-cell conditioned medium (LCM) was generated by filter-sterilizing the supernatant of L-929 cells that were allowed to grow for one week in culture. Primary BMDMs were generated by isolation and culture of mouse bone marrow in complete RPMI supplemented with 20% LCM for up to 12 days. Immortalization of BMDMs was performed as previously described[35] by inoculation with the J2 retrovirus. For cell stimulations, 10 ng/ml LPS (Sigma L8274), 10 ng/ml recombinant human TNF (R&D Systems 210-TA), 100 ng/ml recombinant mouse IL-1β (R&D Systems 401-ML-005), 100 ng/ml recombinant mouse IFNγ (BD Pharmingen 554587), 10 pg/ml recombinant mouse IL-10 (eBioScience 88-7104-ST), 10 μM dexamethasone (Sigma D402) and 0.01 μM oestrogen (Sigma E2758) were used unless otherwise indicated. For tolerization experiments, BMDMs were stimulated with 10 ng/ml LPS for 15 h (or as indicated), washed 5 times with 1× PBS, then allowed to rest for 2 h in LPS-free complete medium supplemented with 20% LCM. BMDMs were then stimulated with 1 μg/ml LPS for 4 h (for qPCR) or 12 h (for enzyme-linked immunosorbent assay (ELISA)), or as indicated.

miRNA microarray. Samples were treated as described, rinsed with 1× PBS, lysed in TRIzol and sent to a commercial microRNA array profiling service (Exiqon). As part of the service, samples were labelled using the miRCURY Hy3/Hy5 Power labelling kit and hybridized on the miRCURY LNA array (v.11.0 hsa, mmu and rno). All capture probes for the control spike-in oligonucleotides produced signals in the expected range. The quantified background-corrected signals were normalized using the global Lowess (locally weighted scatterplot smoothing) regression algorithm, and a list of differentially expressed miRNAs was returned.

miRNA mimic and antagonist oligonucleotides. Pre-miR miRNA precursors (Ambion AM17100) and anti-miR miRNA inhibitors (Ambion AM17000) were transfected into BMDMs to modulate miRNA function in short-term experiments. Part numbers for oligonucleotides are as follows: for overexpression experiments, pre-miR negative control #1 (Invitrogen AM17110), miR-222-3p (PM11376), miR-221-3p (PM10337); for antagonization experiments, anti-miR miRNA negative control #1 (Ambion AM17010), miR-222-3p (AM11376), miR-221-3p (AM10337). To optimize transfection conditions, the FAM dye-labelled pre-miR negative control #1 (Invitrogen AM17121) oligonucleotide was used. Transfection of 50,000 BMDMs per well of a 12-well plate with 6 μl lipofectamine and 0.1 nmol oligonucleotide diluted in 200 μl of Opti-MEM (total) was found to provide transfection of >80% of cells (as measured by flow cytometry), and these conditions were used for all further experiments in BMDMs. Medium was replaced with complete RPMI containing 20% LCM after 4 h to minimize cytotoxicity. Cells were allowed to recover for 24–48 h before stimulation.

Production of virus and BMDM transduction. Plasmids for miRNA overexpression (GeneCopoeia CmiR0001-MR01, MmiR3289-MR01, or MmiR3434-MR01) or antagonization (GeneCopoeia CmiR-AN0001-a.m.03 or HmiR-AN0399-a.m.03) were transfected into 293FT cells with the Lenti-Pac HIV Expression Packaging Kit (GeneCopoeia HPK-LVTR-20) or Lenti-Pac FIV Expression Packaging Kit (GeneCopoeia FPK-LVTR-20) to generate viral particles. BMDMs were inoculated by spin infection in 6-well plates in the presence of 6 μg/ml polybrene (Sigma H9268). Following spin inoculation, viral supernatant was immediately replaced with complete RPMI supplemented with 20% LCM. Cells were allowed to recover overnight. For primary BMDMs, plating for inoculation was generally performed on day five of differentiation. The first spin infection was performed on day six, second spin infection (if necessary) was performed on day seven, and plating for experiments was performed on day eight.

ELISA. BMDMs were plated at 50,000 cells per well, and cytokine concentrations in cell supernatants were measured using the BD OptEIA Mouse IL-6 ELISA Set (BD 555240), BD OptEIA Mouse IL-12 (p40) ELISA Set (BD 555165), or BD OptEIA Mouse TNF (Mono/Mono) ELISA Set (BD 555268) according to the manufacturer's instructions.

RNA extraction, reverse transcription and qPCR. Total RNA was extracted from samples using TRIzol reagent (Invitrogen 15596018). For reverse transcription and detection of miRNAs, the Universal cDNA Synthesis Kit (Exiqon 203301) and locked nucleic acid primers (Exiqon) were used. For other genes, approximately 1 μg of RNA was reverse transcribed with SuperScript III (Invitrogen 18080085). qPCR was then performed with VeriQuest Fast SYBR (Affymetrix 75675). The amplified transcripts were quantified using the comparative C_t method.

Computational prediction of miRNA-binding sites. miR-222-binding sites were predicted using the PITA algorithm[36] (http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html) or MicroCosm Targets program (which utilizes the miRanda algorithm) as indicated in the text. MicroCosm Targets Version 5 was used to search for targets for mmu-miR-222[37]. UTRs and miRNA sequences were manually input to the PITA algorithm, and default search settings were used. All predictions were re-verified with their respective programs on 5 December 2013.

Construction of reporter vectors and luciferase reporter assays. The *Brg1* UTR was amplified from IMAGE clone 30533489 (Open Biosystems MMM1013-9498346) and cloned into the pMIR-Report (Ambion AM5795) multiple cloning site using HindIII and SpeI restriction sites. The *Tnf* UTR was amplified from cDNA generated from BMDMs stimulated with LPS for 1 h, and inserted into the pMIR-Report vector as performed for the *Brg1* UTR. Reporter plasmids were transfected into 293FT cells along with a Renilla luciferase reporter (used to normalize for transfection efficiency). After 24 h, Firefly and Renilla luciferase activity was quantified using the Dual-Luciferase Reporter Assay (Promega E1980).

CRISPR. The CRISPR design tool (http://crispr.mit.edu) was used to design guide RNAs for cloning into the PX458 (Addgene 48138) and PX459 (Addgene 48139) dual Cas9 and single-guide RNA expression plasmids[38] to generate plasmids to target identified miR-222-binding sites for deletion. Cells were transiently transfected with empty vector or targeting vectors. After 24 h, transfected cells were selected using 48 h of puromycin treatment (PX459) or by sorting for GFP positive (PX458) cells. Limiting dilution was performed to isolate clonal cell lines. Clones were screened for appropriate deletion by PCR. Deletion of targeted regions was confirmed by sequencing when necessary. Gene expression was compared between lines with successful deletion, unsuccessful deletion and lines generated by transfection with expression plasmids that lacked a Cas9-targeting sequence.

For deletion of the miR-222-binding site in the *Tnf* UTR, the following guide sequences were used: combination 1, TCAGCGTTATTAAGACAATT GGG and ATTACAGTCACGGCTCCCGT GGG; combination 2, TTGTCTTAATAACGCTGATT TGG and ATTTCTCTCAATGACCCGTA GGG. For deletion of the miR-222-binding site in the *Brg1* UTR, the following guide sequences were used: GGAGTAGCCCTTAGCAGTGA TGG and ACCAGATGTAGTTTCGAACT TGG.

Intracellular staining for flow cytometry. Cells were rinsed and fixed for 15–30 min at room temperature in 4% paraformaldehyde. Cells were rinsed and permeabilized by resuspension in 5% saponin for 10–20 min at room temperature. Either anti-IκBα (L35A5, Cell Signaling 4814), anti-Brg1 (H88, Santa Cruz sc-10768), or rabbit monoclonal antibody IgG isotype control (Cell Signaling 3900) was added, and cells were incubated for an additional 20 min at room temperature. Cells were rinsed and re-suspended in saponin with 1:300 dilution of fluorochrome-conjugated secondary antibody (Alexa Fluor 488 donkey anti-rabbit IgG, Invitrogen A21206; Alexa Fluor 546 goat anti-rabbit IgG, Invitrogen A11010; or Alexa Fluor 546 donkey anti-mouse IgG, Invitrogen A10036). After incubation at room temperature for 20 min, cells were rinsed, re-suspended in PBS and analysed on a BD LSRII flow cytometer.

Chromatin immunoprecipitation. Cells from a 15-cm plate were fixed by incubation in 1% formaldehyde for 5 min, rinsed and lysed by incubation for 5 min on ice in buffer L1 (50 mM Tris at pH 9, 2 mM EDTA, 0.1% NP-40, 10% glycerol, with protease inhibitors). Nuclei were spun down and re-suspended in 500 μl buffer L2 (50 mM Tris at pH 8, 0.1% sodium dodecyl sulphate and 5 mM EDTA). Sonication was performed in a Bioruptor, using 10 cycles of 30 s each. Immunoprecipitation was performed using 20 μl magnetic protein A beads and 5 μg anti-acetyl-histone H4 (Lys5; Millipore 07-327), 2 μg anti-BRG1 (H-88; Santa Cruz sc-10768), or 5 μg anti-acetyl-histone H3 (Millipore 06-599) per 50 μl of chromatin in a 500 μl volume. After overnight rotation at 4 °C, supernatant was isolated. DNA was recovered from the supernatant by adding 20 μl of 5 M NaCl, 50 μl of 10% SDS and 5 μl of proteinase K, shaking for 2 h at 60 °C (unbound fraction). Beads were washed 3× in high salt buffer (20 mM Tris at pH 8.0, 0.1% SDS, 1% NP-40, 2 mM EDTA, and 0.5 M NaCl), and 3× in Tris–EDTA. DNA was eluted from beads by re-suspending beads in 100 μl elution buffer and shaking for 2 h at 60 °C (bound fraction). Bound and unbound fractions were heated to 95 °C for 10 min. DNA was purified from fractions using the Qiagen PCR Purification Kit (28104). To check for promoter binding, qPCR was performed using DNA from the bound and unbound fractions. Bound:unbound ratios were normalized to alpha-crystallin ratios, as this should represent a silent gene.

Amaxa nucleofection. BMDMs were nucleofected with 2 μg of plasmid DNA using the Amaxa Mouse Macrophage Nucleofector Kit (VPA-1009), in conjunction with the Amaxa Nucleofector II Device, according to the manufacturer-optimized protocol.

***Salmonella enterica* serovar Typhimurium infection.** For these experiments, a GFP-expressing *Salmonella enterica* serovar Typhimurium strain (SL1344) was used. *S.* Typhimurium cultures were grown in LB medium supplemented with 100 μg/ml carbenicillin and 30 μg/ml streptomycin. Overnight cultures were diluted and allowed to grow for an additional hour before use, to ensure bacteria were in log-growth phase. $OD_{600\,nm}$ readings were correlated to previously determined CFU values and used to quantify number of bacteria present in culture. BMDMs were infected by inoculation of DMEM growth medium (containing only streptomycin) with bacteria at a multiplicity of infection of 50. Plates were spun at 800 r.c.f. for

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

5 min at 4 °C. BMDMs were incubated for 30 min at 37 °C. Cells were washed 3 times, and then incubated in medium containing gentamycin (100 μg/ml for incubations of 2 h or less, 12 μg/ml for longer incubations). BMDMs were subsequently analysed for GFP content by flow cytometry, or lysed in water to enable plating of lysate dilutions on LB agar plates containing carbenicillin to determine bacterial CFU counts.

Mice. For BMDM generation, female C57Bl/6J mice of 7–10 weeks of age were used, unless otherwise noted. For tolerance and septic shock experiments, male C57Bl/6J mice of 6–10 weeks of age were used. LPS (*Escherichia coli* O55:B5; Sigma L2880) and D-(+)-galactosamine hydrochloride (Sigma G0500) were re-suspended in sterile PBS and filter-sterilized before intraperitoneal injection. For in vivo infection experiments, mice were given intraperitoneal injections of 1×10^7 CFUs per kg of a GFP-expressing *Salmonella enterica* serovar Typhimurium strain (SL1344) suspended in PBS. Mice were maintained under specific-pathogen-free conditions in animal facilities at Columbia University Medical Center. All animal experiments were carried out with the approval of the Columbia University Institutional Animal Care and Use Committee, and in compliance with regulations and guidelines set forth by Columbia University.

Generation of knockout mice. *mir-221 mir-222* knockout mice were generated at the Columbia University Transgenic Mouse facility. In brief, KV1 (129B6 hybrid) ES cells were electroporated with the linearized targeting construct discussed in Extended Data Fig. 6. After positive and negative selection, clonal cell lines were screened by PCR for proper integration of the construct. Positive lines were expanded, blastocyst injection was performed and germline transmission was confirmed. *mir-221 mir-222* knockout mice were backcrossed to the C57Bl/6 background 5–8 times before experimental use.

Peritoneal macrophage isolation. Five millilitres of cold PBS was injected into the peritoneal cavity of euthanized mice. The peritoneum was gently massaged. Fluid was collected and the process was repeated. Cell suspension was spun down and cells were plated at 500,000 cells per well in 12-well plates. Macrophages were allowed to adhere overnight. Non-adherent cells were rinsed off with PBS washes.

Thioglycollate elicitation of peritoneal macrophages. Three per cent thioglycollate was sterilized and aged for at least two months. One millilitre of thioglycolate preparation was injected into the peritoneal cavity of each mouse five days before the isolation of macrophages (as described in 'Peritoneal macrophage isolation').

Monocyte isolation. Bones were isolated from wild-type C57Bl6/J mice. Marrow was retrieved by crushing. Monocytes were purified using the EasySep Mouse Monocyte Isolation Kit.

RNA sequencing. RNA sequencing was performed by the JP Sulzberger Columbia Genome Center. Poly-A pull-down was used to enrich mRNAs from total RNA samples (200 ng–1 μg per sample, RIN > 8 required). Libraries were prepared using the Illumina TruSeq RNA prep kit. Libraries were then sequenced using Illumina HiSeq2000. Multiplexed and pooled samples were sequenced to a depth of 24–34 $\times 10^6$ reads per sample as 100-bp single-end reads. RTA (Illumina) was used for base calling, and bcl2fastq (version 1.8.4) was used for converting BCL to fastq format, coupled with adaptor trimming. Reads were mapped to a reference genome (mouse: UCSC/mm9) using Tophat (version 2.1.0) with 4 mismatches (–read-mismatches = 4) and 10 maximum multiple hits (–max-multihits = 10). To tackle the mapping issue of reads that are from exon–exon junctions, Tophat infers novel exon–exon junctions ab initio and combines them with junctions from known mRNA sequences (refgenes) as the reference annotation. The relative abundance (also known as the expression level) of genes and splice isoforms were estimated using cufflinks (version 2.0.2) with default settings.

ChIP–seq analysis. Track data of genes of interest were loaded into Galaxy[39] (http://usegalaxy.org) using the UCSC table browser and mouse mm10 genome. Using Galaxy, previously published ChIP–seq data[20–25] was then aligned to the mouse mm10 genome using the HISAT program (Galaxy version 2.03) with default settings. BamCoverage (Galaxy version 2.3.6.0) was then used to generate a coverage bigwig file, using default settings to scale to the size of the mm9 mouse genome. ComputeMatrix (Galaxy version 2.3.6.0) and plotHeatmap (Galaxy version 2.3.6.0) were then used to compare transcription factor occupancy at gene promoters, using the transcription start site as the reference point.

Dataset references. ChIP–seq data that were analysed were from the European Nucleotide Archive accession ERA319838 (IRF5), and from the following Gene Expression Omnibus accessions: GSE56123[20] (IRF1, IRF8, STAT1, STAT2); GSE67343[21] (IRF3); GSE36104[22] (IRF2, IRF4, NF-κB subunits); GSE62697[23] (IRF7); GSE77886[24] (IRF mutants); and GSE38379[25] (STAT1 knockout).

Patient sample selection and processing. We selected 10 consecutive patients, newly admitted to a medical or surgical intensive care unit (ICU), who had systemic inflammatory response syndrome and a known or suspected infection[40]. Patients were excluded from the study if they had an ICU admission or bacteraemia within the previous 30 days. After obtaining informed consent from the patient or a surrogate, whole blood was drawn within 4 h of ICU admission. Peripheral blood mononuclear cells were isolated from whole blood of healthy human volunteers or buffy coat isolates from patients from the ICU meeting sepsis criteria, by centrifugation on a Ficoll cushion. RNA was isolated with the miRNeasy micro kit (Qiagen 217084) and reverse transcribed as in 'RNA extraction, reverse transcription and qPCR'. Experiments were performed with approval of the Institutional Review Board at Columbia University and in accordance with regulations and guidelines set forth by the university. The results of these experiments are shown in Fig. 4a.

Patient sample selection and processing. Additional patient cohorts were obtained from hospitalized patients with acute decompensation of chronic liver disease and suspected bacterial infection. Baseline characteristics and outcome of patients with decompensated liver disease in the absence or presence of multiple organ failure syndrome (according to the EASL CLIF-C criteria for acute-on-chronic liver failure[41]) are given in Extended Data Fig. 8. Clinical scores—such as model for end-stage liver disease scores, bacterial culture count, protein analysis, blood count and serum levels of C-reactive protein and creatinine—were obtained from routine laboratory analysis. The determination of serum concentration of TNF was performed by ELISA. The results of these experiments are shown in Fig. 4b–f.

The isolation and characterization of human immune cells and the use of clinical data was approved by the internal review board (ethics committee of the Jena University Hospital, no. 3683-02/3). The study conformed to the ethical guidelines of the 1975 Declaration of Helsinki, and patients granted written informed consent before inclusion.

Statistics and sample collection. Student's *t*-tests were performed using the T.TEST function in Microsoft Excel. All other statistical tests were performed using Prism software. Unless otherwise stated, two-sided tests were performed. For samples using cell lines and cells isolated from inbred mice, the Student's *t*-test was often used. The distributional requirements for the test are assumptions. This means that—for instance—under the assumption of normal-distributed residuals, the *t*-test is an exact test. However, given a non-normal distribution of cell line data the test is not exact, but is instead approximative. For patient samples, nonparametric tests were used to avoid the assumption of a normal distribution. In all figures, error bars represent s.e.m. unless otherwise indicated. Standard deviations and s.e.m. values were calculated for each group of data and used to estimate variation (s.e.m. values are shown as error bars in most experiments). Variation generally appears similar between groups being compared. All experiments were replicated in the laboratory at least two times. Unless otherwise indicated, in experiments using primary cells *n* represents the number of experiments performed with separate cell isolations; in experiments using immortalized cells or cell lines, *n* represents the number of experiments performed using separate cell populations. Systematic randomization and blinding were not performed. Samples were excluded from the analysis if they were identified as outliers using the Grubbs' test (also known as the extreme studentized deviate method).

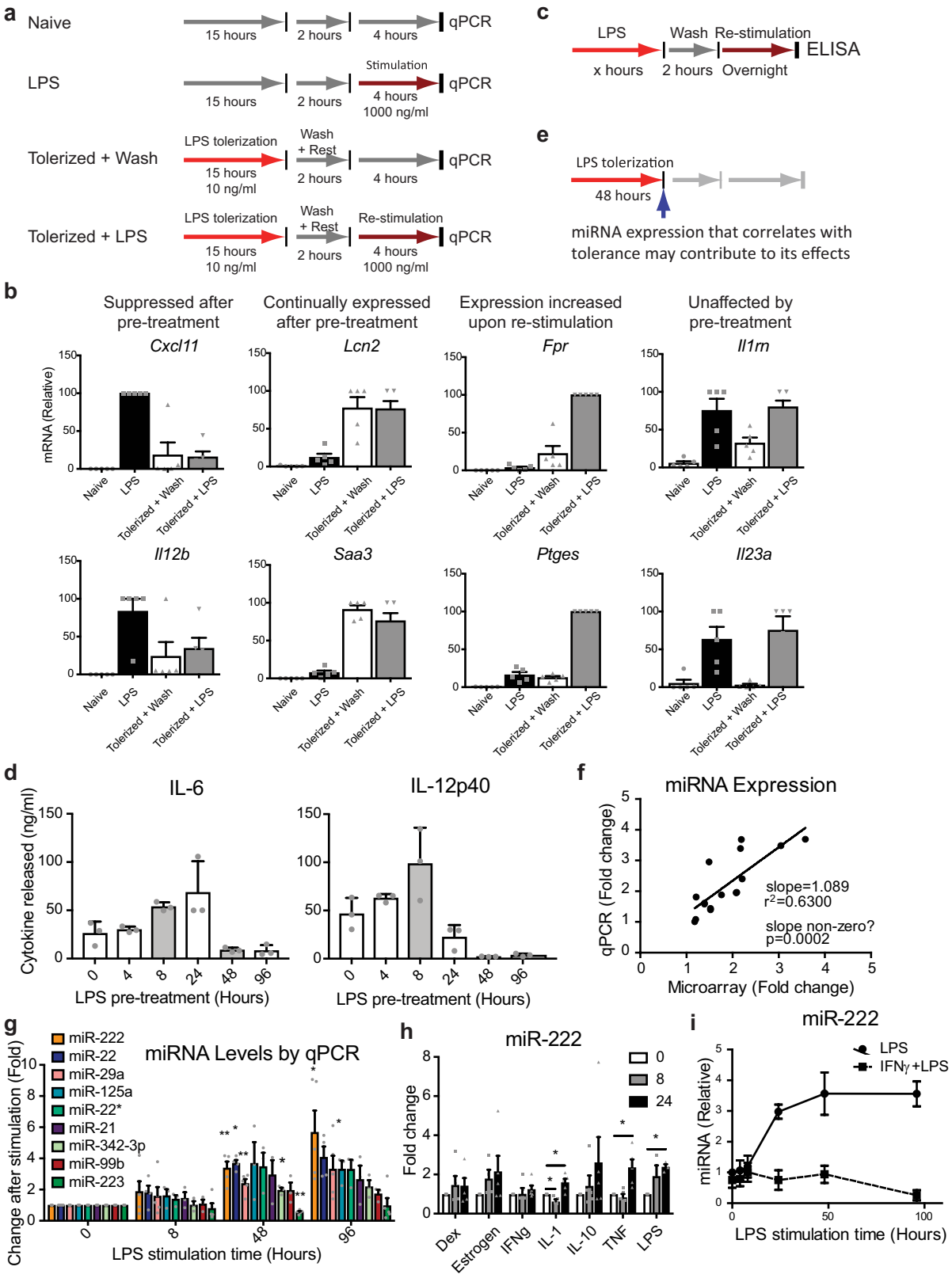
For mouse LPS-shock studies, an appropriate sample size was estimated on the basis of an outcome variable of survival time, measured in hours. An estimate was based on using a one-tailed Student's *t*-test to determine statistical significance. Control mice were expected to succumb within 62 h. Knockout mice were expected to become moribund 52 h after LPS injection at the latest. Therefore, the minimal effect size was estimated to be 10 h. On the basis of the literature and experiments previously performed by our laboratory, we anticipated a standard deviation of 10 h. Taking into account a power of 80% and alpha of 0.05, we calculated a sample size of 10 mice per genotype.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data accessibility. RNA sequencing data that support the findings of this study have been deposited in Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) with the accession code GSE89918. All other data are available from the corresponding author upon reasonable request.

35. Roberson, S. M. & Walker, W. S. Immortalization of cloned mouse splenic macrophages with a retrovirus containing the v-*raf/mil* and v-*myc* oncogenes. *Cell. Immunol.* **116**, 341–351 (1988).
36. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nat. Genet.* **39**, 1278–1284 (2007).
37. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34**, D140–D144 (2006).
38. Ran, F. A. et al. Genome engineering using the CRISPR–Cas9 system. *Nat. Protocols* **8**, 2281–2308 (2013).
39. Afgan, E. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44**, W3–W10 (2016).
40. Bone, R. C. et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Chest* **101**, 1644–1655 (1992).
41. Moreau, R. et al. Acute-on-chronic liver failure is a distinct syndrome that develops in patients with acute decompensation of cirrhosis. *Gastroenterology* **144**, 1426–1437.e9 (2013).
42. Rao, P. et al. IκBβ acts to inhibit and activate gene expression during the inflammatory response. *Nature* **466**, 1115–1119 (2010).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

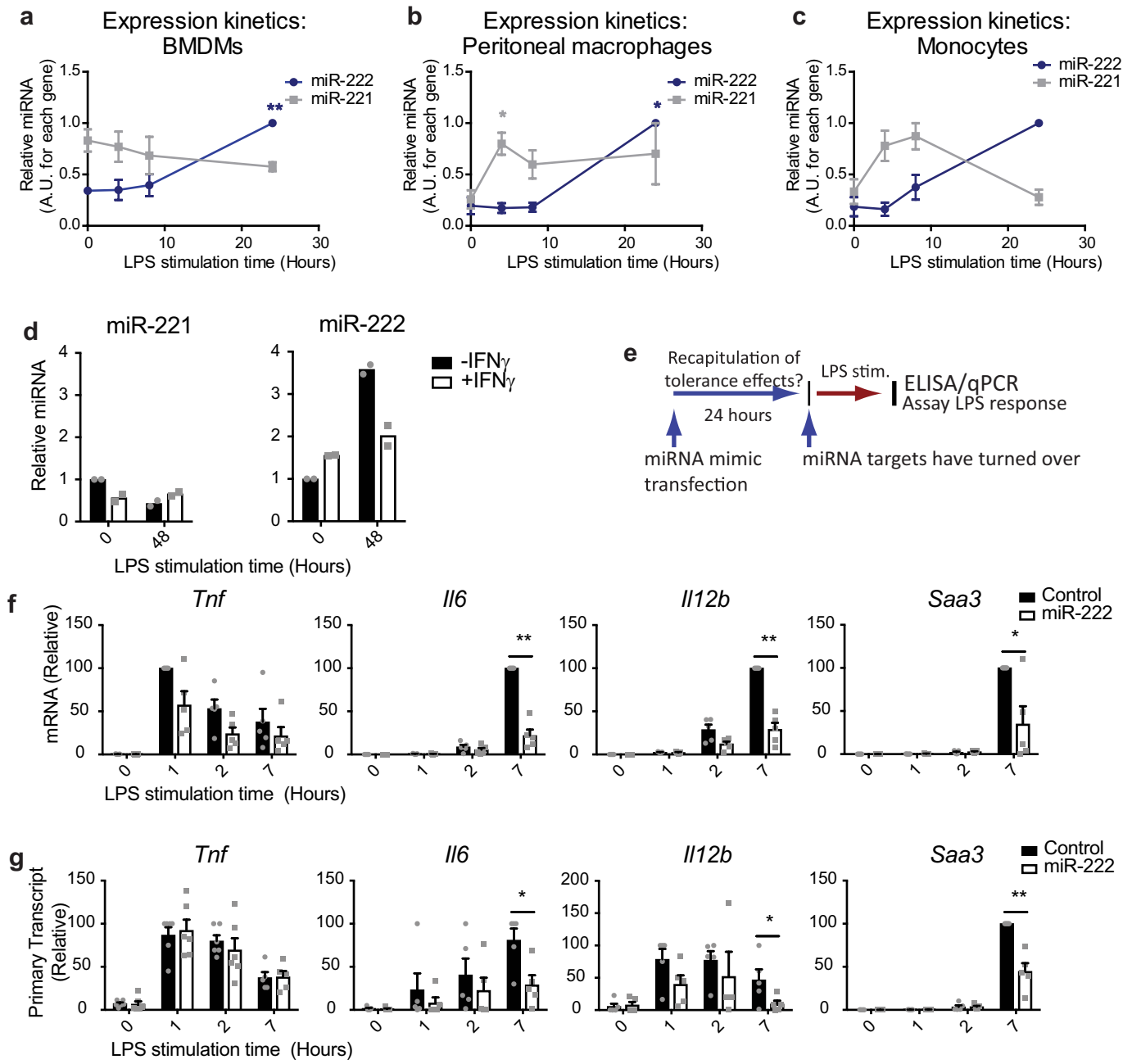


Extended Data Fig. 1 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

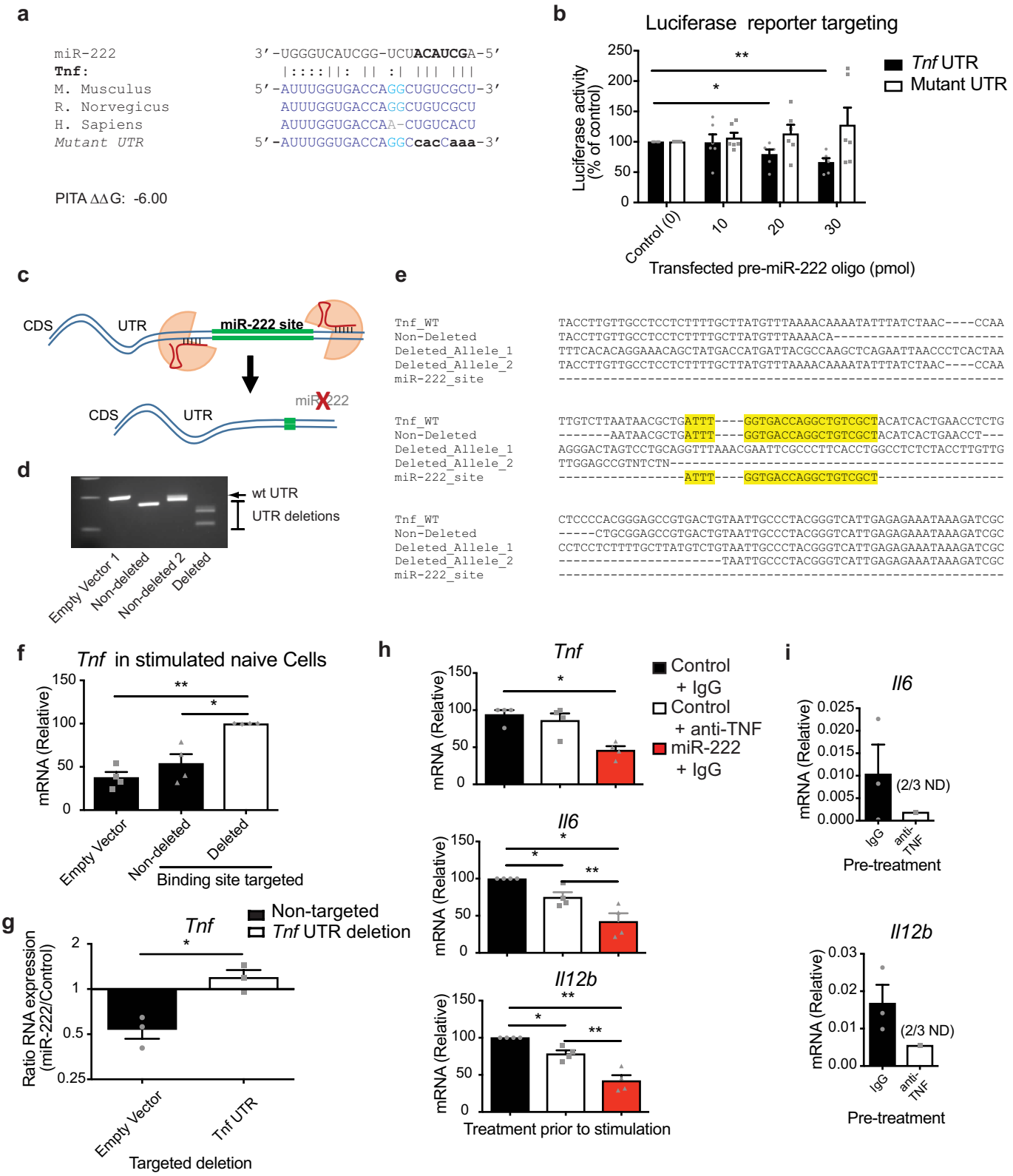
Extended Data Fig. 1 | In vitro modelling of tolerance and miR-222 induction upon prolonged LPS stimulation. **a**, Schematic of experiments performed in **b**. **b**, Expression of LPS-response genes in control BMDMs that have undergone the given treatments. Four major expression patterns of LPS-response genes in response to tolerization were noted ($n = 5$ biologically independent samples). **c**, Schematic of experiments performed in **d**. **d**, Cytokine production—measured by ELISA—by BMDMs re-stimulated with LPS overnight after pre-treatment with LPS for the given periods of time. Time points chosen for miRNA microarray analysis have bars shaded in grey ($n = 3$ biologically independent samples). **e**, Schematic of strategy for experiments performed in Fig. 1. **f**, Comparison of microarray (x axis) and qPCR (y axis) measurements of LPS-induced upregulation of miRNAs. A linear regression showing the correlation between the two methods is plotted ($n = 16$ miRNAs tested). **g**, qPCR verification of LPS-induced change in expression of nine miRNAs ($n = 3$ biologically independent samples). **h**, Expression of miR-222 after stimulation of BMDMs by anti-inflammatory and tolerance-inducing factors for the given lengths of time ($n = 5$ biologically independent samples; Dex, dexamethasone). **i**, Expression of miR-222 in response to LPS alone, or LPS after pre-treatment of BMDMs with IFNγ ($n = 4$ biologically independent samples). For all bar and line graphs, mean ± s.e.m. is plotted. $^{*}P < 0.05$, $^{**}P < 0.01$, determined by two-sided Student's t-test for paired values.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 2 | Differential regulation of miR-222 and miR-221 and association of miR-222 with in vitro tolerance. a–c, Expression of miR-221 and miR-222 in response to LPS stimulation of BMDMs (**a**, $n = 4$ biologically independent samples), peritoneal macrophages (**b**, $n = 3$ biologically independent samples for miR-222 and $n = 4$ biologically independent samples for miR-221) or monocytes isolated from the bone marrow (**c**, $n = 3$ biologically independent samples), as determined by qPCR. **d**, LPS-induced miR-221 and miR-222 expression in BMDMs with or without IFNγ pre-treatment, as determined by qPCR ($n = 2$ biologically independent samples). **e**, Schematic of experiments performed in **f**, **g** and Fig. 1c. **f**, **g**, LPS-induced gene expression at the mRNA (**f**) or primary transcript (**g**) level after miR-222 mimic transfection ($n = 5$ biologically independent samples). For all bar and line graphs, mean ± s.e.m. is plotted. $^{*}P < 0.05$, $^{**}P < 0.01$, determined by two-sided Student's *t*-test for paired values.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

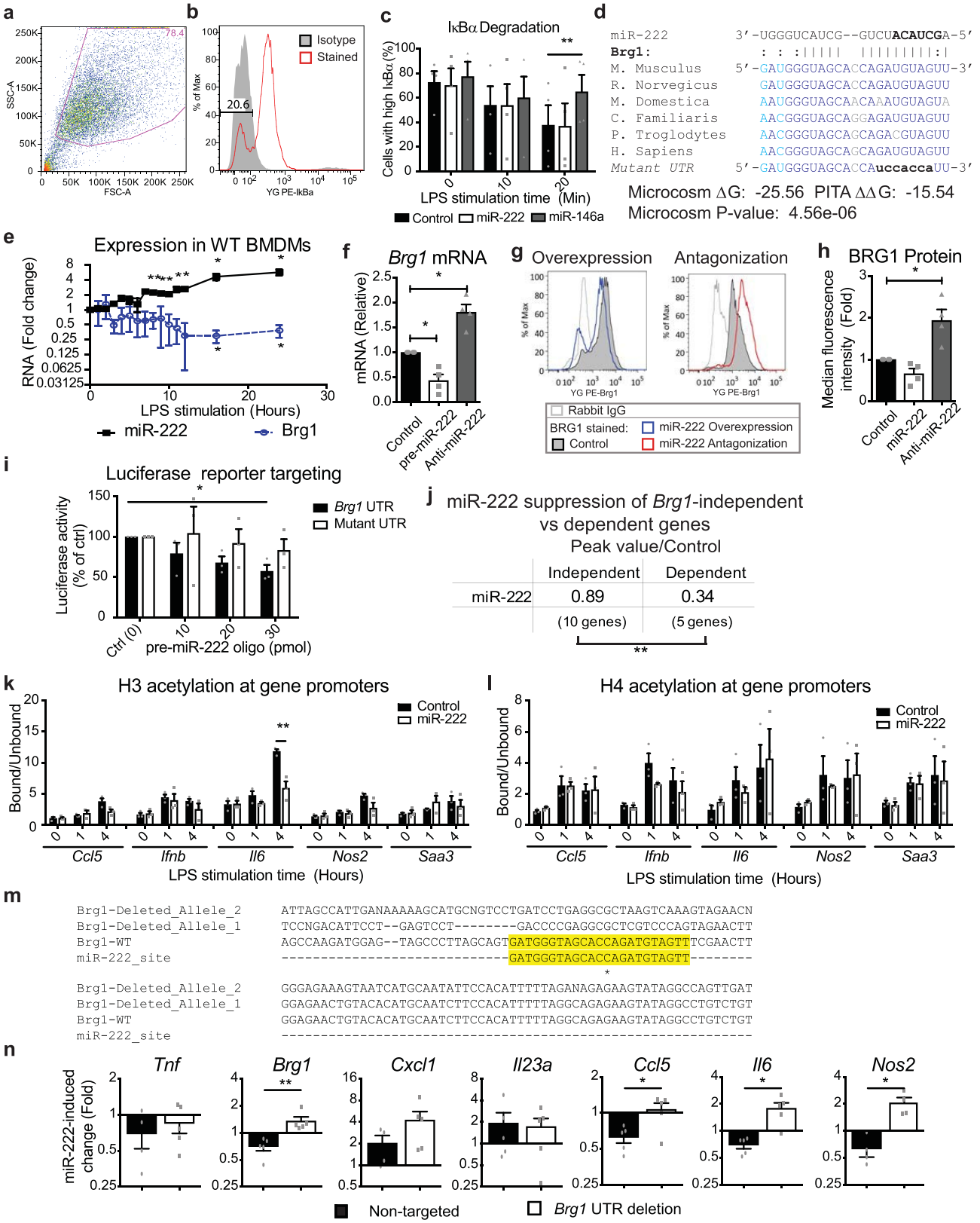


Extended Data Fig. 3 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 3 | *Tnf* is a direct target of miR-222, but suppression of *Tnf* does not account for miR-222-mediated transcriptional silencing of late LPS-response genes. **a**, Sequence and prediction scores of an miR-222-binding site in the *Tnf* UTR. **b**, Activity of a luciferase reporter construct in which the luciferase coding sequence is followed by either the complete *Tnf* UTR, or a UTR in which the predicted miR-222-binding site has been mutated to the sequence shown in **a** ($n = 6$ independent experiments). **c**, CRISPR–Cas9 targeting strategy to delete predicted binding sites. CDS, coding sequence. **d**, Clones of RAW cells were screened for successful deletion of the miR-222-binding site by PCR across the targeted region of the UTR, using genomic DNA from the given clonal line as a template. Screening for *Tnf* UTR deletion is shown. Experiment was repeated twice with similar results. **e**, Successful deletion of the miR-222-binding site in RAW cell clones was confirmed by sequencing genomic DNA of the given cell line. miR-222-binding site in the *Tnf* UTR is highlighted in yellow. **f**, LPS-induced *Tnf* expression in control and CRISPR–Cas9-targeted RAW cells ($n = 4$ independent experiments). **g**, Average effect of miR-222 mimic transfection on LPS-induced *Tnf* mRNA levels in either control mouse embryonic fibroblasts or mouse embryonic fibroblasts that have undergone CRISPR targeting and clonal selection for deletion of the miR-222-binding site. Average of the effects from the three clonal lines ($n = 3$ independent experiments) is shown. **h**, Wild-type BMDMs were transfected with a control or miR-222 mimic oligonucleotide. Twenty-four hours later, cells were pre-treated with an isotype control (IgG) or TNF-neutralizing (anti-TNF) antibody for two hours, and stimulated with 10 ng ml^{-1} LPS. Expression of the given genes was measured by qPCR ($n = 4$ biologically independent samples). **i**, Efficacy of TNF neutralization was confirmed by treating cells with IgG or anti-TNF as above, followed by stimulation with 100 ng ml^{-1} recombinant mouse TNF ($n = 3$ biologically independent samples). Gene upregulation was not detected (ND) in two out of three samples treated with anti-TNF. For all bar graphs, mean ± s.e.m. is plotted. *$P < 0.05$, **$P < 0.01$, determined by two-sided Student's *t*-test for paired values.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

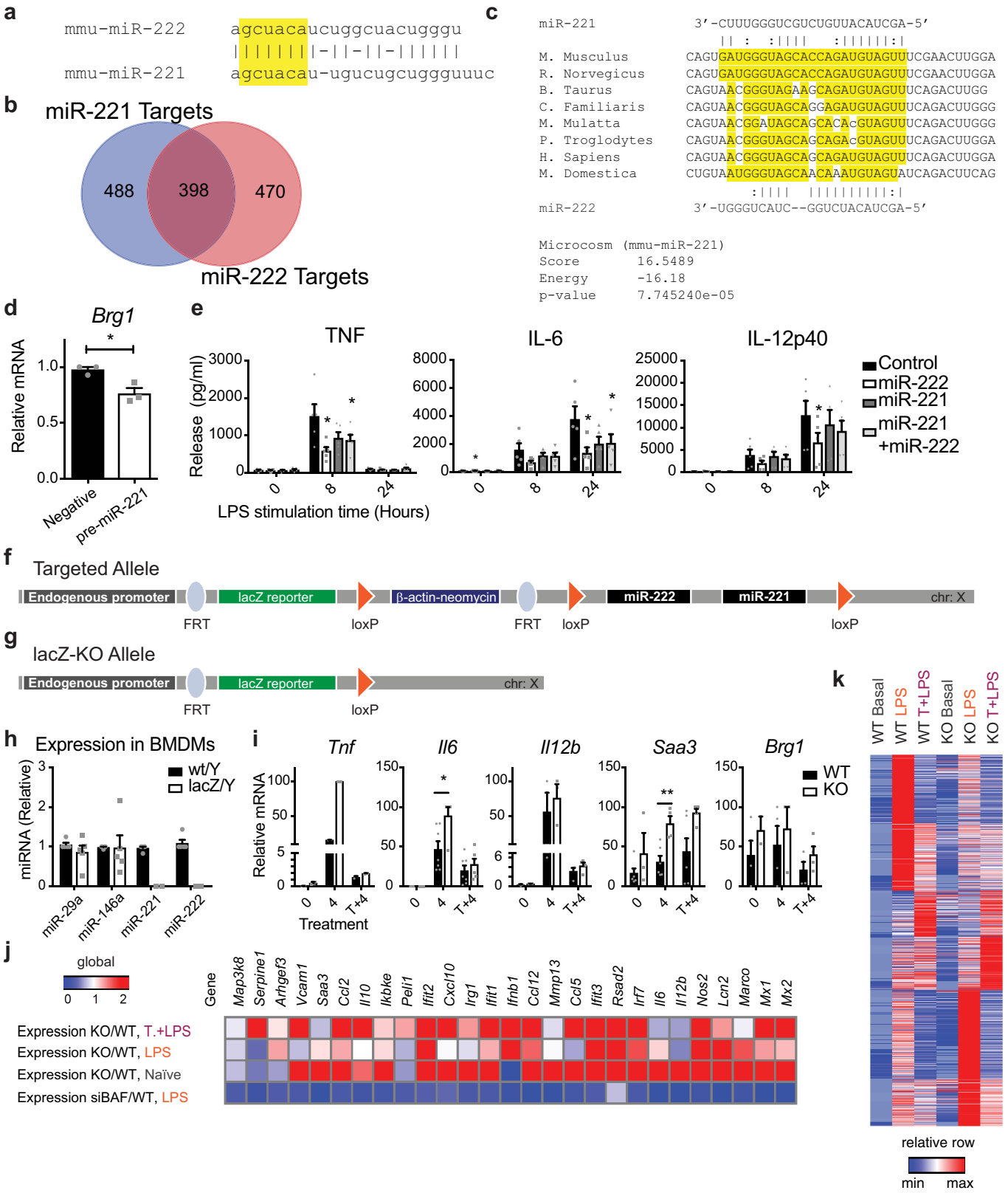


Extended Data Fig. 4 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 4 | Evidence of miR-222 targeting of *Brg1*. **a**, Example of gating used to exclude dead cells from flow cytometry analyses in **c**, **g** and Extended Data Fig. 6i. **b**, Example of gating used to distinguish cells with high versus low levels of IκBα, as analysed in **c**. **c**, Effect of miRNA overexpression (by viral transduction) on LPS-induced IκBα degradation in immortalized BMDMs, measured by flow cytometry ($n = 4$ independent experiments). **d**, Sequence and prediction scores of an miR-222-binding site in the *Brg1* UTR. **e**, miR-222 and *Brg1* mRNA levels in LPS-stimulated BMDMs ($n = 3$ biologically independent samples). **f**, *Brg1* mRNA levels in resting BMDMs 24 h after transfection ($n = 4$ biologically independent samples). **g**, Effect of miRNA overexpression or antagonization (by viral transduction) on BRG1 levels in immortalized BMDMs, observed by flow cytometry. Representative of four independent experiments with similar results, quantified in **h**. **h**, Flow cytometry analysis of BRG1 protein levels in transduced immortalized BMDMs ($n = 4$ independent experiments). **i**, Activity of a luciferase reporter construct in which the luciferase coding sequence is followed by either the complete *Brg1* UTR, or a UTR in which the predicted miR-222-binding site has been mutated to the sequence shown in **d** ($n = 3$ independent experiments). **j**, Quantification of the average effect of miR-222 mimic transfection on *Brg1*-dependent and *Brg1*-independent LPS-response genes ($n = 3$ biologically independent samples). Two-sided Student's *t*-test for heteroscedastic values used to compare ratios (ratio of miR-222 overexpression to control) at peak LPS-induced expression times for *Brg1*-dependent versus *Brg1*-independent genes. **k**, **l**, ChIP for histone H3 acetylation (**k**) or histone H4 acetylation (**l**) after LPS stimulation of immortalized BMDMs transduced with overexpression constructs (**k** and **l** tested in same $n = 3$ independent experiments). **m**, Successful deletion of the miR-222-binding site in the *Brg1* UTR in RAW cell clones was confirmed by sequencing genomic DNA of the given cell line. miR-222-binding site is highlighted in yellow. **n**, Effect of miR-222 overexpression (by oligonucleotide transfection) on LPS-induced gene expression in either a RAW cell line in which the *Brg1*–miR-222 binding site was deleted by CRISPR targeting (as shown in Extended Data Fig. 3c) or a cell line in which the binding site was not targeted for deletion ($n = 5$ independent experiments). For all bar graphs, mean ± s.e.m. is plotted. $^{*}P < 0.05$, $^{**}P < 0.01$, determined by two-sided Student's *t*-test for paired values.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 5 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 5 | Comparison of miR-221 and miR-222, and effects of miR-221 and miR-222 deletion on the transcriptional response to LPS. a, Alignment of the mature miR-221 and miR-222 sequences. The miRNA seed sequence is highlighted in yellow. **b**, Venn diagram displaying overlap between MicroCosm target predictions for mmu-miR-221 and mmu-miR-222. **c**, Alignment and computational scores of miR-221 sequence with predicted *Brg1* UTR target site. Alignment of miR-222 sequence with the site is also shown. **d**, *Brg1* expression in BMDMs transfected with the given oligonucleotide ($n = 3$ biologically independent samples). **e**, LPS-induced cytokine production in BMDMs transfected with given miRNA mimics, as measured by ELISA ($n = 5$ biologically independent samples). **f**, Schematic of the miR-221 and miR-222 locus after targeting with a construct designed to generate both complete and conditional *mir-221 mir-222* knockout mice. **g**, Schematic of the miR-221 and miR-222 locus after breeding targeted mice (**f**) with EIIa-Cre mice, which results in complete deletion of miR-221 and miR-222. **h**, miRNA expression in BMDMs from littermates with a wild-type or *mir-221 mir-222* knockout allele ($n = 5$ biologically independent samples). **i**, LPS-induced gene expression in naive or tolerized peritoneal macrophages isolated from wild-type or *mir-221 mir-222* knockout littermates ($n = 7$ biologically independent samples). **j**, Heat map comparing the effect of *Brg1* and *Brm* knockdown[15] and *mir-222* knockout on gene expression. Colours represent values of the given ratios; red indicates increased expression, white indicates no change and blue indicates decreased expression. **k**, Heat map of LPS-induced gene expression in wild-type and *mir-221 mir-222* knockout macrophages. For all bar graphs, mean ± s.e.m. is plotted. $*P < 0.05$, $**P < 0.01$, determined by two-sided Student's *t*-test for paired (**d**, **e**) or heteroscedastic (**i**) values.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

a Genes higher basally in KO

PANTHER GO-Slim Biological Process	Fold Enrichment	+/-	P value
response to interferon-gamma	10.63	+	1.66E-09
↳immune response	3.41	+	5.37E-08
↳immune system process	2.41	+	8.95E-03
cytokine-mediated signaling pathway	8.78	+	4.81E-24
localization	8.15	+	8.14E-13
locomotion	7.16	+	1.38E-13

Genes lower basally in KO

PANTHER GO-Slim Biological Process	Fold Enrichment	+/-	P value
macrophage activation	4.13	+	2.43E-02
cell-cell adhesion	2.98	+	1.98E-03
signal transduction	2.46	+	3.05E-03
Unclassified	.79	-	0.00E00

b Genes higher in LPS-stimulated KO

PANTHER GO-Slim Biological Process	Fold Enrichment	+/-	P value
response to interferon-gamma	15.14	+	5.61E-03
↳immune response	5.47	+	2.23E-04
cytokine-mediated signaling pathway	10.24	+	3.54E-06
macrophage activation	8.92	+	1.59E-02
cell-cell signaling	7.72	+	9.69E-03

Genes lower in LPS-stimulated KO

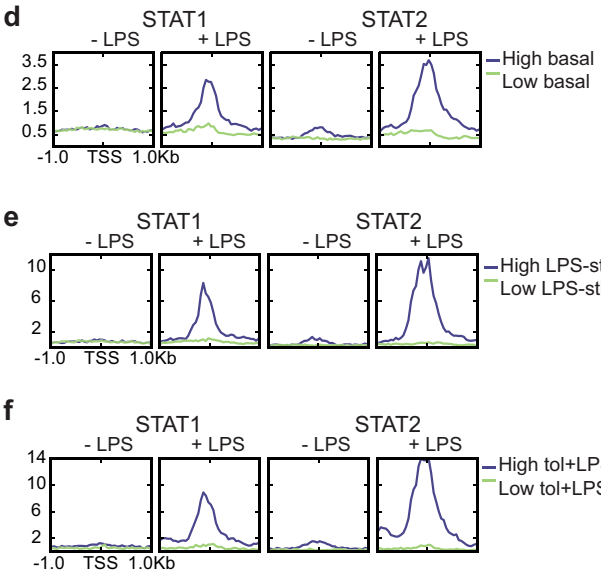
PANTHER GO-Slim Biological Process	Fold Enrichment	+/-	P value
Unclassified	.67	-	0.00E00

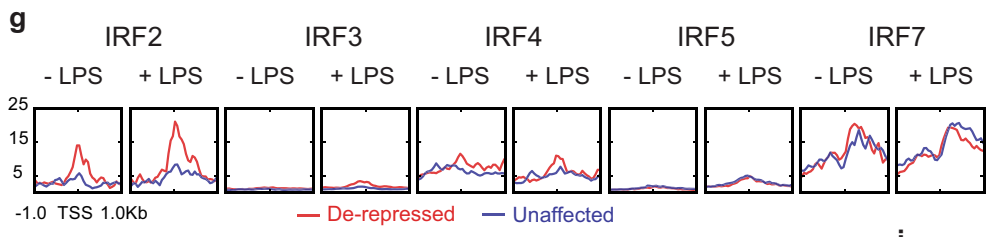
c Genes higher in tolerized+LPS-stimulated KO

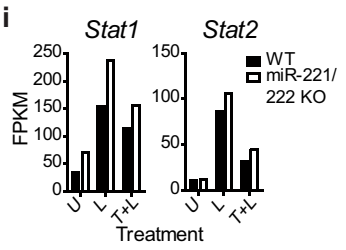
PANTHER GO-Slim Biological Process	Fold Enrichment	+/-	P value
response to interferon-gamma	28.99	+	1.24E-07
↳immune response	5.54	+	1.34E-03
cytokine-mediated signaling pathway	17.81	+	3.24E-13
localization	12.46	+	4.41E-04
locomotion	9.43	+	2.68E-03

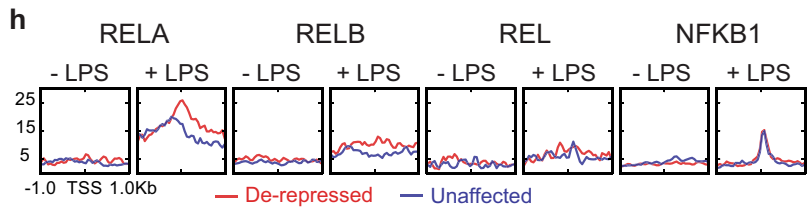
Genes lower in tolerized+LPS-stimulated KO

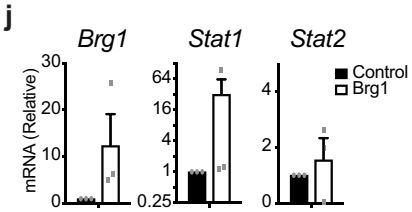
PANTHER GO-Slim Biological Process	Fold Enrichment	+/-	P value
behavior	32.35	+	2.80E-02
localization	22.14	+	7.92E-04
locomotion	16.76	+	3.02E-03
cytokine-mediated signaling pathway	16.63	+	3.99E-04





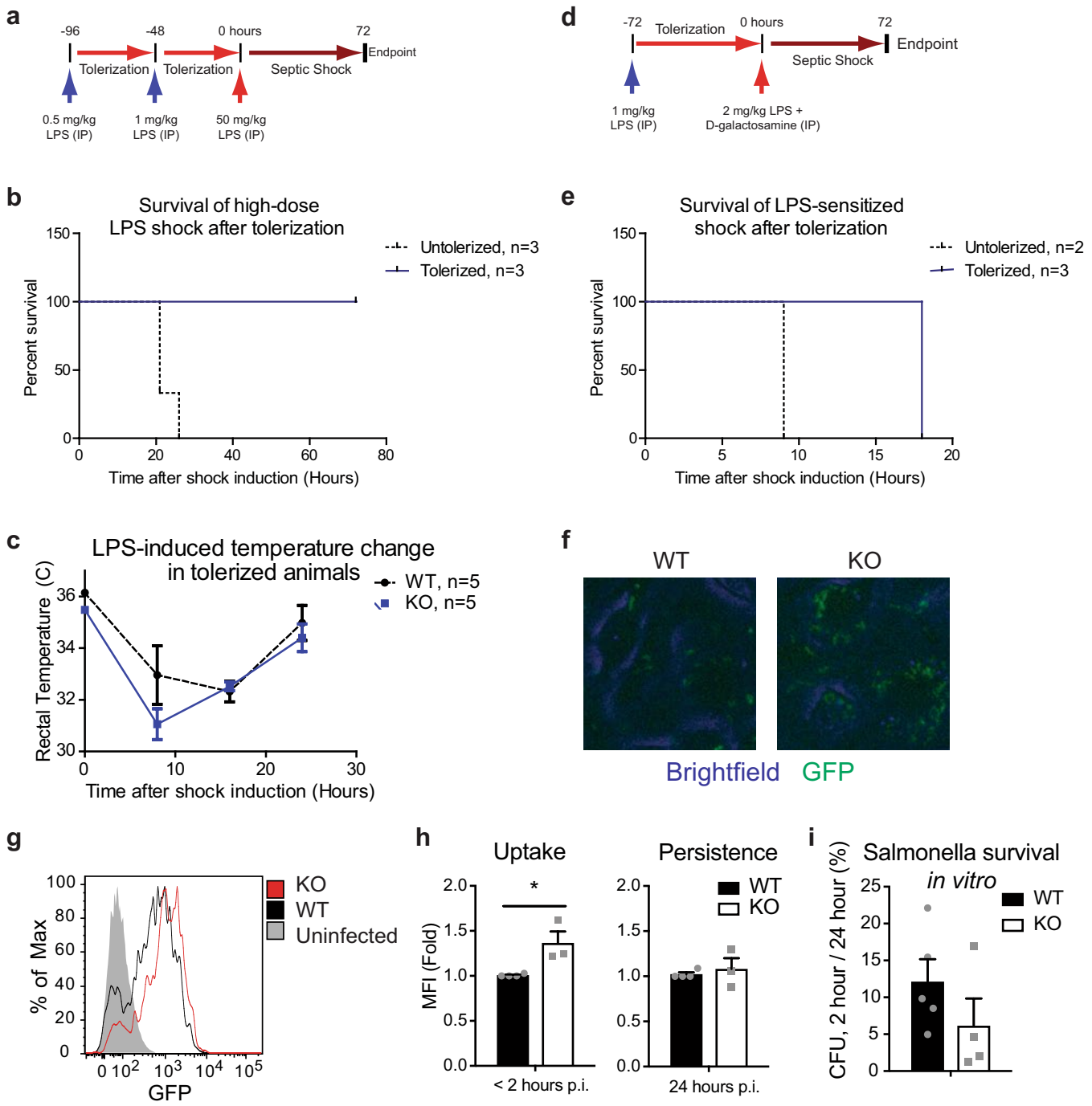






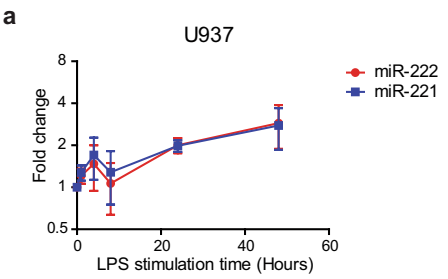
Extended Data Fig. 6 | Gene Ontology and ChIP–seq analysis shows that genes affected by *mir-221 mir-222* knockout have differential gene functions and transcription-factor binding at promoters. a–f, Enriched GO terms (**a**–**c**) and transcription-factor binding at promoters (**d**–**f**) of genes that are expressed at higher (twofold or higher) or lower (0.5-fold or lower) levels in *mir-221 mir-222* knockout macrophages after no stimulation (**a**, **d**, $n = 647$ genes higher, 565 genes lower), LPS stimulation (**b**, **e**, $n = 143$ genes higher, 121 genes lower) or LPS tolerization followed by re-stimulation (**c**, **f**, $n = 123$ genes higher,48 genes lower). PANTHER was used to identify GO terms. The top four terms for each category are shown; GO terms that are unique to either higher- or lower-expression gene subsets are highlighted. **g**, **h**, IRF and NF-κB subunit occupancy at gene promoters; gene subsets analysed are described in Fig. 2h. For transcription factor analyses, previously published ChIP–seq data were used[20–25]. **i**, RNA levels of genes in wild-type or *mir-221 mir-222* knockout peritoneal macrophages, quantified by a single RNA sequencing experiment. **j**, qPCR for gene expression in wild-type BMDMs after Amaxa-based nucleofection of given overexpression construct ($n = 3$ biologically independent samples). For all bar graphs, centre value represents the mean and errors bars (if applicable) represent s.e.m.

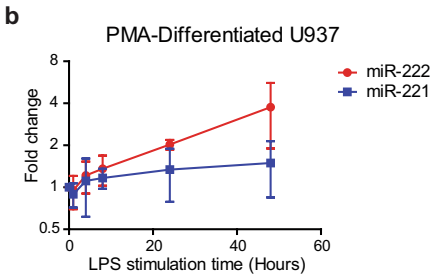
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 7 | *mir-221 mir-222* knockout mice have an altered LPS response and knockout macrophages exhibit enhanced *Salmonella* uptake and clearance in vitro. a, Schematic of experiments performed in **b**, **c**. **b**, Survival of naive or tolerized mice injected with high doses of LPS. **c**, Wild-type or *mir-222* knockout littermates were tolerized to LPS before lethal LPS injection. The change in body temperature after final LPS injection was monitored for 24 h. **d**, Schematic of experiments performed in **e**. **e**, Survival of naive or tolerized mice injected with LPS and D-galactosamine. **f**, BMDMs from wild-type or *mir-221 mir-222* knockout mice were spin-infected with a GFP-expressing strain of *S.* Typhimurium. Fluorescence was analysed by microscopy 60 min after infection. Representative of two independent experiments with similar results. **g**, BMDMs from wild-type or *mir-221 mir-222* knockout mice were spin-infected with a GFP-expressing strain of *S.* Typhimurium. Fluorescence was analysed by flow cytometry 30 min post-infection. Representative of three independent experiments with similar results. **h**, Average fluorescence of infected BMDMs after early (left) or late (right) time points after infection ($n = 4$ biologically independent wild-type samples, 3 biologically independent knockout samples). **i**, Survival of *S.* Typhimurium after in vitro infection of BMDMs, determined by comparing CFUs after lysis of BMDMs at early and late time points of infection ($n = 5$ biologically independent wild-type samples, 4 biologically independent knockout samples). For all bar and line graphs, mean ± s.e.m. is plotted. $^{*}P < 0.05$, $^{**}P < 0.01$, determined by two-sided Student's *t*-test for heteroscedastic values.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.





c

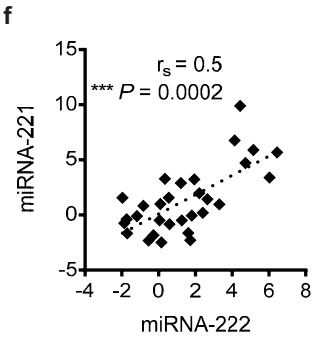
Characteristic	
Sex	
Male	6 (60%)
Female	4 (40%)
Age (years)	63 (53-74)
Operative reason for admission	
Yes	3 (30%)
No	7 (70%)
Mechanical ventilation	
Yes	5 (50%)
No	5 (50%)
White blood cell count (cells/mm3 x 1000)	12.9 (7.7-17.9)
Albumin (g/L)	2.5 (2.2-3.4)
Urine output (mL/24h)	895 (579-2463)
Pre-ICU length of stay (days)	1 (0-4)
APACHE IV score	62 (48-113)
SAPS II	48 (38-97)

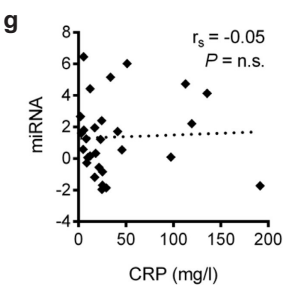
d

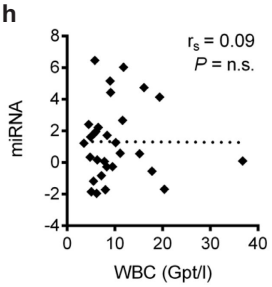
	Acute Decompensation without organ failure N=16	Acute Decompensation with organ failure (ACLF) N=14	P value
Age (years)	58 (50-63)	64 (49-72)	0.29
Male sex	14 (88%)	11 (79%)	0.64
Alcoholic liver disease	11 (69%)	12 (86%)	0.40
ACLF grade (I/II/III)	0	9 (64%) / 3 (21%) / 2 (14%)	N/A
Bacterial infection			
- None	12 (75%)	8 (57%)	0.44*
- Peritonitis	2 (13%)	4 (29%)	
- Urinary tract	2 (13%)	1 (7%)	
- Pneumonia	0	1 (7%)	
Creatinine (µmol/l)	89 (75-108)	233 (186-299)	<0.001
Total bilirubin (µmol/l)	18 (11-47)	115 (25-310)	0.003
MELD score	10 (9-19)	29 (24-34)	<0.001
White blood cells ($x10^3$/µl)	6.9 (5.7-10.6)	9.3 (5.6-16.9)	0.28
C-reactive protein (mg/l)	18 (9-28)	25 (10-105)	0.19
Liver transplant or death within 30 days**	1 (6%)	6 (43%)	0.03

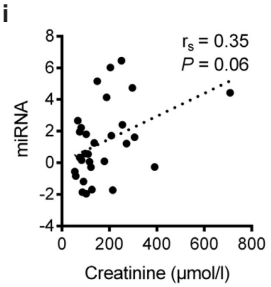
e

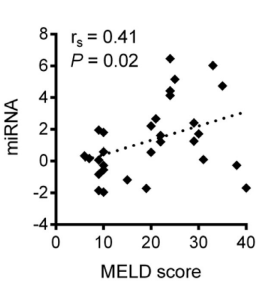
	Acute Decompensation without organ failure N=5	Acute Decompensation with organ failure (ACLF) N=5	P value
Age (years)	59 (53-70)	54 (46-59)	0.20
Male sex	5 (100%)	4 (80%)	1.00
Alcoholic liver disease	4 (80%)	5 (100%)	1.00
ACLF grade (I/II/III)	0	4 (80%) / 0 (0%) / 1 (20%)	N/A
Bacterial infection			
- None	5 (100%)	0	0.008*
- Peritonitis	0	5 (100%)	
- Urinary tract	0	0	
- Pneumonia	0	0	
Creatinine (µmol/l)	92 (52-110)	193 (123-367)	0.008
Total bilirubin (µmol/l)	16 (11-23)	57 (43-146)	0.008
MELD score	10 (9-11)	28 (21-36)	0.008
White blood cells ($x10^3$/µl)	7.2 (5.3-9.6)	13.6 (6.0-19.7)	0.22
C-reactive protein (mg/l)	33 (28-84)	76 (27-146)	0.75
Liver transplant or death within 30 days**	0	4 (80%)	0.048









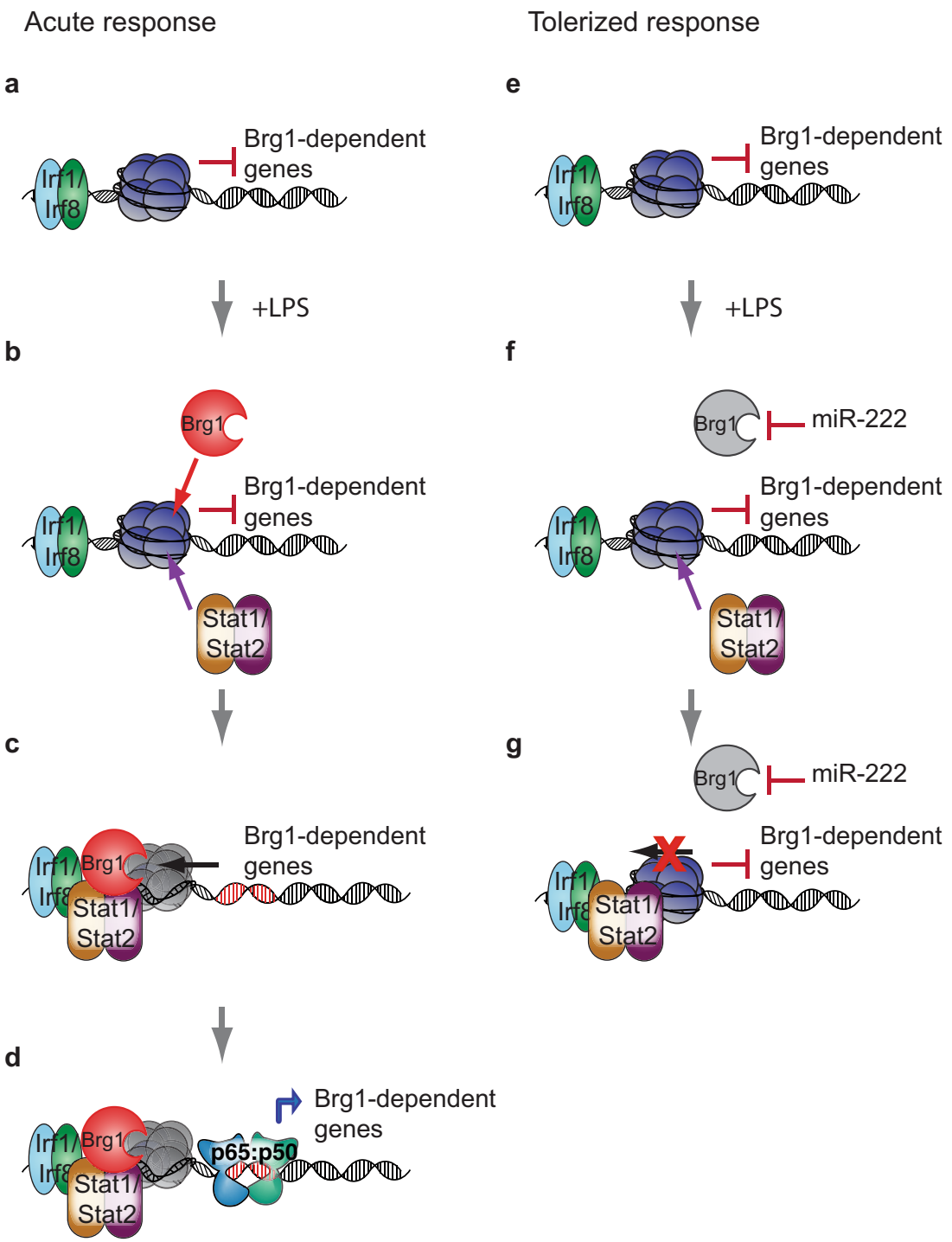


Extended Data Fig. 8 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 8 | miR-221 and miR-222 are upregulated in human cells and patients with sepsis. a, **b**, LPS-induced miRNA expression in undifferentiated (**a**) or phorbol 12-myristate 13-acetate-differentiated (**b**) human U937 cells ($n = 3$ independent experiments). **c**, Patient characteristics for data shown in Fig. 4a. APACHE, acute physiology and chronic health evaluation; SAPS, simplified acute physiology score. Categorical variables are given as *n* (percentage in parentheses) and continuous variables as median (interquartile range in parentheses). **d**, **e**, Baseline characteristics of patients with decompensated liver disease in the absence or presence of multiple organ failure syndrome (according to the EASL CLIF-C criteria for acute-on-chronic liver failure). Data in **d** correspond to peripheral blood mononuclear cell analyses (Fig. 4b–d). Median with interquartiles or frequencies and percentages are shown. *P* values from Mann–Whitney *U* test or Fisher's exact test as appropriate (two-sided). *, comparing any infection to no infection; **, 4 out of 30 (13%) and 1 out of 10 (10%) patients were lost to follow-up within 30 days. Data in **e** correspond to monocyte analyses (Fig. 4f). Median with interquartiles or frequencies and percentages are shown. *P* values from Mann–Whitney *U* test or Fisher's exact test as appropriate (two-sided). *, comparing any infection to no infection; **, 1 out of 10 (10%) patients was lost to follow-up within 30 days. **f**, Correlation between miR-221 and miR-222 levels in patients characterized in **d** ($n = 30$ patients). Bivariate nonparametric correlation analysis (Spearman's rho) was used to identify correlations between variables and *P* values. **g**–**j**, Linear correlation of miR-222 expression and C-reactive protein (**g**), white blood cell count (**h**), creatinine levels (**i**) or model for end-stage liver disease score (**j**) in samples from the patient cohort described in **d** ($n = 30$ patients). Bivariate nonparametric correlation analysis (Spearman's rho) was used to identify correlations between variables and *P* values. For line graphs, mean ± s.e.m. is plotted.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 9 | Model of the effect of miR-222 on LPS-induced macrophage tolerance. a, Before an acute LPS stimulation, chromatin at BRG1-dependent gene promoters prevents binding of remodelling-dependent transcription factors and RNA polymerase. **b**, After an acute LPS stimulation, transcription factors such as STAT1 and STAT2 are recruited to gene promoters and stabilize BRG1 binding. **c**, **d**, BRG1 activity leads to chromatin remodelling (**c**), which enables recruitment of additional transcription factors, such as NF-κB, to the unwound DNA (**d**). This enables polymerase recruitment and licensing, which leads to gene transcription. **e**, After an initial LPS response, chromatin is 'reset' to an inhibitory state by negative regulators of chromatin accessibility. **f**, Upon LPS re-stimulation, transcription factors must again be recruited to gene promoters. However, miR-222 limits the level of BRG1. **g**, Lack of available BRG1 prevents chromatin remodelling at many gene promoters, and prevents downstream transcription factor recruitment. This prevents gene transcription from occurring in most cells.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Identification of targets of miR-222

Predicted Target	Algorithm Score	P-Value	% Decrease
Mesdc1	16.1968	3.86E-09	31.83
Nfyb	16.5309	8.58E-07	31.90
Nfyb	16.027	8.58E-07	31.90
Sntb1	15.4316	2.95E-06	25.70
Smarca4	17.4905	4.57E-06	22.64
Dclre1a	15.2548	8.64E-06	22.67
Nudt5	16.5017	2.21E-05	82.49
Tpbg	16.3439	3.79E-05	75.81
Ptx3	15.9272	4.03E-05	50.04
Apaf1	15.3191	9.59E-05	39.79
Atp1a1	17.6386	9.74E-05	50.18
Pdhb	15.579	1.55E-04	26.18
Uchl1	15.4257	4.72E-04	20.26
Dhx9	16.4603	8.57E-04	96.82
Tsc2	20.1182	8.93E-04	40.97
Stmn1	16.5573	1.11E-03	82.55
Stmn1	16.009	1.11E-03	82.55
Ogfr	16.0031	1.13E-03	21.35
Ogfr	16.0031	1.13E-03	21.35
Ddx52	15.8837	1.30E-03	22.54
Zfp462	15.3834	1.55E-03	20.92
Sap30	17.1462	2.13E-03	37.05
Mad2l2	16.0031	2.22E-03	37.34
Idh2	15.8888	2.87E-03	47.13
Il19	17.0841	3.61E-03	53.02
Slc28a1	15.7179	4.02E-03	97.33
Tsc2	18.4316	4.61E-03	40.97
Capn7	15.7361	4.74E-03	24.89
Aldh2	15.5459	5.13E-03	23.55
Agpat2	18.2641	5.42E-03	40.86
Kcnh2	16.7713	5.47E-03	32.31
Cdca3	15.5835	5.53E-03	47.93
Plaur	15.7368	5.98E-03	57.51
Agpat2	18.1381	6.13E-03	40.86
Nfkbil1	17.832	6.74E-03	22.93
Slc23a3	17.9463	7.38E-03	96.59
Zyx	15.3325	7.52E-03	27.10
Nudt12	15.3316	7.53E-03	47.39
Nfkb1	15.676	7.54E-03	67.70
Nnt	15.7463	7.66E-03	30.32

(list continues on right)

Predicted Target	Algorithm Score	P-Value	% Decrease
Lcp1	16.0977	7.72E-03	45.12
Lrg1	15.3173	7.73E-03	22.69
Grip1	17.8508	8.09E-03	22.34
Golga1	15.2869	8.15E-03	28.15
Mapk6	15.3013	9.89E-03	36.07
Smarca4	17.4953	1.14E-02	22.64
Camp	15.6238	1.43E-02	39.31
Slc25a11	15.3352	1.46E-02	59.66
Sult1a1	15.2729	1.60E-02	31.32
4930544G11Rik	17.1192	1.64E-02	24.04
Cish	17.0804	1.70E-02	26.38
Pdcd10	15.3459	1.88E-02	63.49
Slc23a3	16.9058	2.01E-02	96.59
Qdpr	16.4511	2.06E-02	50.84
Pabpc1	16.86	2.10E-02	65.55
Cacnb2	16.0474	2.12E-02	34.44
Ddhd1	16.8345	2.15E-02	89.58
Dbnl	16.7678	2.29E-02	29.83
Ddhd1	16.696	2.45E-02	89.58
Rtn1	16.6889	2.47E-02	33.46
Exosc5	16.6889	2.47E-02	22.56
Fignl1	16.64	2.59E-02	23.59
2610020O08Rik	16.527	2.88E-02	21.06
Tnfsf11	16.4951	2.97E-02	22.56
Atox1	16.4786	3.02E-02	21.37
Fntb	16.3904	3.28E-02	26.90
4933402D24Rik	16.3904	3.28E-02	21.47
Olfr110	16.3782	3.32E-02	30.81
Mrpl3	15.4861	3.34E-02	63.03
Azi2	16.3295	3.47E-02	27.88
Tnks	16.326	3.48E-02	25.84
Fkbp9	15.872	3.58E-02	60.69
S100a4	16.2317	3.80E-02	78.00
Dtymk	16.2317	3.80E-02	20.44
Ppp1r14d	16.1174	3.82E-02	80.53
H47	16.2002	3.92E-02	24.67
Mrg2	16.1772	4.00E-02	32.25
Rpl27a	16.0727	4.41E-02	24.60
Impa2	16.0513	4.50E-02	66.38
Smarca4	16.0312	4.59E-02	22.64
Mad2l2	16.0031	4.71E-02	37.34
Cd33	15.9762	4.83E-02	46.63

miR-222 targets predicted by the MicroCosm program were filtered on the basis of their expression in macrophages. Only targets that decreased in expression after between 8 and 24 h of LPS stimulation (column 4) were considered (using microarray data generated in a previous study[42]). Results were then sorted by *P* value (generated by the MicroCosm program). *Brg1* is highlighted in red. Note that multiple listings for a target indicate that more than one site prediction for that gene was made by the MicroCosm program.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

https://doi.org/10.1038/s41586-018-0256-2

OTULIN limits cell death and inflammation by deubiquitinating LUBAC

Klaus Heger[1], Katherine E. Wickliffe[1], Ada Ndoja[1], Juan Zhang[2], Aditya Murthy[3], Debra L. Dugger[1], Allie Maltzman[1], Felipe de Sousa e Melo[4], Jeffrey Hung[5], Yi Zeng[6], Erik Verschueren[6], Donald S. Kirkpatrick[6], Domagoj Vucic[7], Wyne P. Lee[2], Merone Roose-Girma[8], Robert J. Newman[8], Søren Warming[8], Yi-Chun Hsiao[9], László G. Kőműves[5], Joshua D. Webster[5], Kim Newton[1]* & Vishva M. Dixit[1]*

OTULIN (OTU deubiquitinase with linear linkage specificity) removes linear polyubiquitin from proteins that have been modified by LUBAC (linear ubiquitin chain assembly complex) and is critical for preventing auto-inflammatory disease[1,2] and embryonic lethality during mouse development[3]. Here we show that OTULIN promotes rather than counteracts LUBAC activity by preventing its auto-ubiquitination with linear polyubiquitin. Thus, knock-in mice that express catalytically inactive OTULIN, either constitutively or selectively in endothelial cells, resembled LUBAC-deficient mice[4] and died midgestation as a result of cell death mediated by TNFR1 (tumour necrosis factor receptor 1) and the kinase activity of RIPK1 (receptor-interacting protein kinase 1). Inactivation of OTULIN in adult mice also caused pro-inflammatory cell death. Accordingly, embryonic lethality and adult auto-inflammation were prevented by the combined loss of cell death mediators: caspase 8 for apoptosis and RIPK3 for necroptosis. Unexpectedly, OTULIN mutant mice that lacked caspase 8 and RIPK3 died in the perinatal period, exhibiting enhanced production of type I interferon that was dependent on RIPK1. Collectively, our results indicate that OTULIN and LUBAC function in a linear pathway, and highlight a previously unrecognized interaction between linear ubiquitination, regulators of cell death, and induction of type I interferon.

Mouse embryos carrying the OTULIN(W96R) or OTULIN(D336E) mutations have vascularization defects that have been attributed to impaired Wnt signalling[3], whereas OTULIN loss in mouse haematopoietic cells or homozygous hypomorphic *OTULIN* mutations in humans cause severe inflammation that has been attributed to enhanced NF-κB-dependent gene transcription[1,2]. It has been proposed that OTULIN counteracts LUBAC activity within TNFR1, NOD2 (nucleotide binding oligomerization domain containing 2), and Wnt signalling complexes[3,5–8]; however, even though LUBAC and OTULIN form a stable complex[3,5,7,9], only LUBAC is recruited to TNFR1 and NOD2 signalling complexes[10,11]. Therefore, the key substrates of OTULIN and where OTULIN acts to limit inflammation remain unclear.

To determine the physiological function of OTULIN enzymatic activity, we analysed *Otulin*$^{C129A/C129A}$ knock-in mice expressing the active site cysteine mutant OTULIN(C129A) (Extended Data Fig. 1a). While *Otulin*$^{C129A/+}$ mice were viable (Extended Data Fig. 1b), *Otulin*$^{C129A/C129A}$ embryos were resorbed after exhibiting abnormal yolk sac vasculature at embryonic day 10.5 (E10.5) (Fig. 1a, Extended Data Fig. 1c–e). Staining for cleaved caspase 3, a marker of apoptosis, or RIPK3 autophosphorylated on Thr231 and Ser232, a marker of necroptosis, revealed extensive cell death in both the yolk sac and placenta, with some dying cells having endothelial morphology (Fig. 1b, Extended Data Fig. 1f–i). Linear polyubiquitin was also increased in E10.5 *Otulin*$^{C129A/C129A}$ embryos (Extended Data Fig. 1j), consistent with OTULIN being the main deubiquitinase for linear polyubiquitin[3,5].

Restricting expression of OTULIN(C129A) to endothelial cells and some haematopoietic cells using an inducible *Otulin*^{iC129A} allele (Extended Data Fig. 1k) and a *Cdh5-Cre* transgene[12] also caused embryonic lethality (Extended Data Fig. 1l). Disruption of the yolk sac vasculature at E11.5 (Fig. 1c, Extended Data Fig. 1m) was preceded by extensive apoptosis at E10.5 (Fig. 1d, Extended Data Fig. 1n), consistent with OTULIN activity promoting endothelial cell survival during embryogenesis.

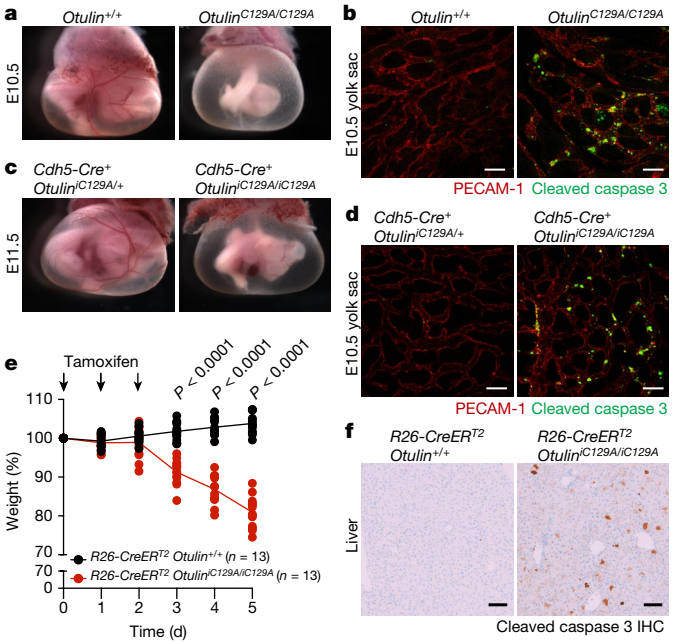


Fig. 1 | OTULIN inactivation causes embryonic lethality and weight loss. a, E10.5 embryos ($n = 15$ *Otulin*$^{+/+}$, $n = 23$ *Otulin*$^{C129A/C129A}$). **b**, **d**, E10.5 yolk sacs stained for PECAM-1 (red) and cleaved caspase 3 (green) ($n = 5$ *Otulin*$^{+/+}$, $n = 9$ *Otulin*$^{C129A/C129A}$ in **b**; $n = 3$ *Cdh5-Cre*$^{+}$ *Otulin*$^{iC129A/+}$, $n = 4$ *Cdh5-Cre*$^{+}$ *Otulin*$^{iC129A/iC129A}$ in **d**). Scale bar, 50 μm. **c**, E11.5 embryos ($n = 4$ *Cdh5-Cre*$^{+}$ *Otulin*$^{iC129A/+}$, $n = 3$ *Cdh5-Cre*$^{+}$ *Otulin*$^{iC129A/iC129A}$). **e**, Body weight of tamoxifen-treated mice plotted as a percentage of initial body weight. Lines represent the mean. *P* values, repeated measures two-way ANOVA followed by Bonferroni's multiple comparison test. **f**, Livers from day 5 in **e** stained for cleaved caspase 3 (brown) ($n = 5$ per genotype). Scale bar, 100 μm.

[1]Department of Physiological Chemistry, Genentech, South San Francisco, CA, USA. [2]Department of Translational Immunology, Genentech, South San Francisco, CA, USA. [3]Department of Cancer Immunology, Genentech, South San Francisco, CA, USA. [4]Department of Molecular Oncology, Genentech, South San Francisco, CA, USA. [5]Department of Pathology, Genentech, South San Francisco, CA, USA. [6]Department of Microchemistry, Proteomics and Lipidomics, Genentech, South San Francisco, CA, USA. [7]Department of Early Discovery Biochemistry, Genentech, South San Francisco, CA, USA. [8]Department of Molecular Biology, Genentech, South San Francisco, CA, USA. [9]Department of Antibody Engineering, Genentech, South San Francisco, CA, USA. *e-mail: knewton@gene.com; dixit@gene.com

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

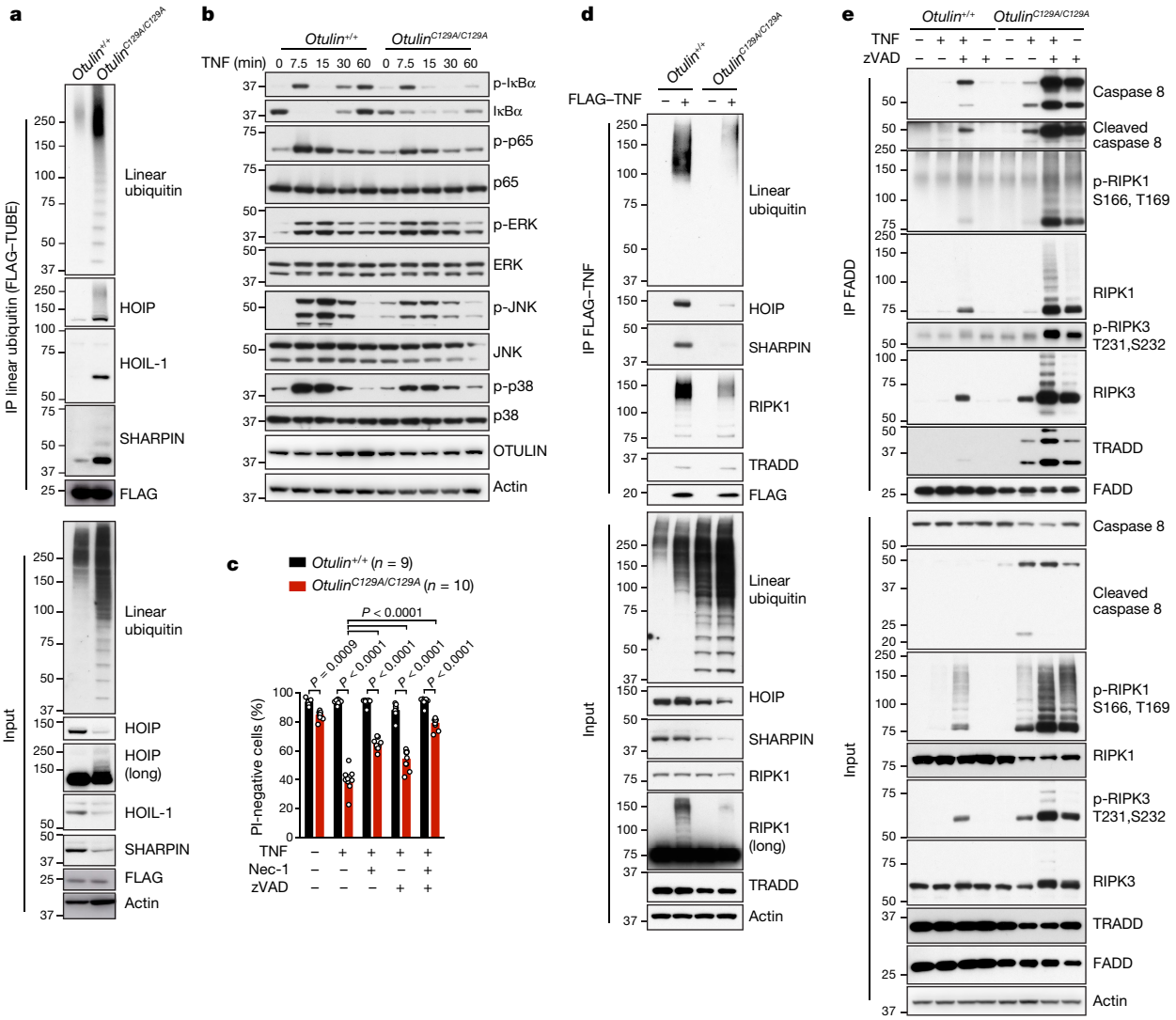


Fig. 2 | OTULIN counteracts LUBAC auto-ubiquitination to suppress cell death. a, **b**, **d**, **e**, Western blots of MEFs or their immunoprecipitates (IP). Treatments were for 5 min (**d**) or 4 h (**e**). Results are representative of two (**a**, **e**) or three (**b**, **d**) independent experiments. A FLAG-tagged TUBE (tandem ubiquitin binding entity) reagent was used to pull-down linear polyubiquitin. **c**, Graph indicates the percentage of MEFs that were viable and not stained by propidium iodide (PI) after overnight treatment. Bars represent the mean. Circles represent MEFs from individual mice. *P* values, one-way ANOVA followed by Tukey's multiple comparison test.

Inactivation of OTULIN in adult mice using a tamoxifen-inducible *Rosa26* (*R26*)-*CreER*^{T2} allele[13] caused rapid weight loss (Fig. 1e). Neutrophils and pro-inflammatory cytokines and chemokines were elevated in the blood, and there was increased cell death and tissue degeneration in the bone marrow, thymus, liver, small intestine, and heart (Fig. 1f, Extended Data Fig. 2). Radiation-induced bone marrow chimaeras with tamoxifen-induced OTULIN(C129A) expression restricted to the haematopoietic compartment also exhibited weight loss, elevated pro-inflammatory cytokines and chemokines, and immune cell deregulation, but they lacked extensive cell death in the liver and small intestine (Extended Data Fig. 3). Therefore, OTULIN activity is important in non-haematopoietic as well as haematopoietic cells. Indeed, when *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ mice reconstituted with wild-type bone marrow were treated with tamoxifen, they showed mild weight loss, mildly elevated cytokines, and degeneration in the liver and small intestine (Extended Data Fig. 4).

Similar to *Otulin*$^{C129A/C129A}$ embryos (Extended Data Fig. 1j), primary mouse embryo fibroblasts (MEFs) and immortalized haematopoietic progenitor cells expressing OTULIN(C129A) contained abundant linear polyubiquitin (Extended Data Fig. 5a–c). This polyubiquitin was protein-conjugated because it was not cleaved by the deubiquitinase IsoT (Extended Data Fig. 5d). Mass spectrometry of the proteins modified with linear polyubiquitin in *Otulin*$^{C129A/C129A}$ cells identified the core LUBAC components HOIP, HOIL-1 and SHARPIN (Extended Data Fig. 5e), which is consistent with regulation of LUBAC auto-ubiquitination by OTULIN[6,10].

Enhanced linear ubiquitination of LUBAC in *Otulin*$^{C129A/C129A}$ MEFs coincided with a reduction in unmodified HOIP, HOIL-1 and SHARPIN (Fig. 2a). A similar reduction in HOIP and SHARPIN has been reported for OTULIN-deficient lymphocytes[1]. LUBAC disappearance required its E3 ligase activity because SHARPIN, HOIL-1 and HOIP were maintained in fibroblasts expressing catalytically inactive HOIP(C879S) and OTULIN(C129A) (Extended Data Fig. 5f, g). Note that *Otulin*$^{C129A/C129A}$ lysates treated with recombinant OTULIN or the promiscuous deubiquitinase USP2 still lacked LUBAC, despite successful elimination of the linear polyubiquitin signal (Extended Data Fig. 5h). Therefore, the drop in unmodified LUBAC did not simply reflect the retarded migration of increasingly ubiquitinated LUBAC. Both transcriptional and post-translational mechanisms may contribute to the decrease in LUBAC because *Hoip* (also known as *Rnf31*), *Hoil-1* (also known as *Rbck1*) and *Sharpin* mRNAs were reduced slightly in *Otulin*$^{C129A/C129A}$ cells (Extended Data Fig. 5i–k), and proteasome inhibition caused a subtle increase in LUBAC components (Extended

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

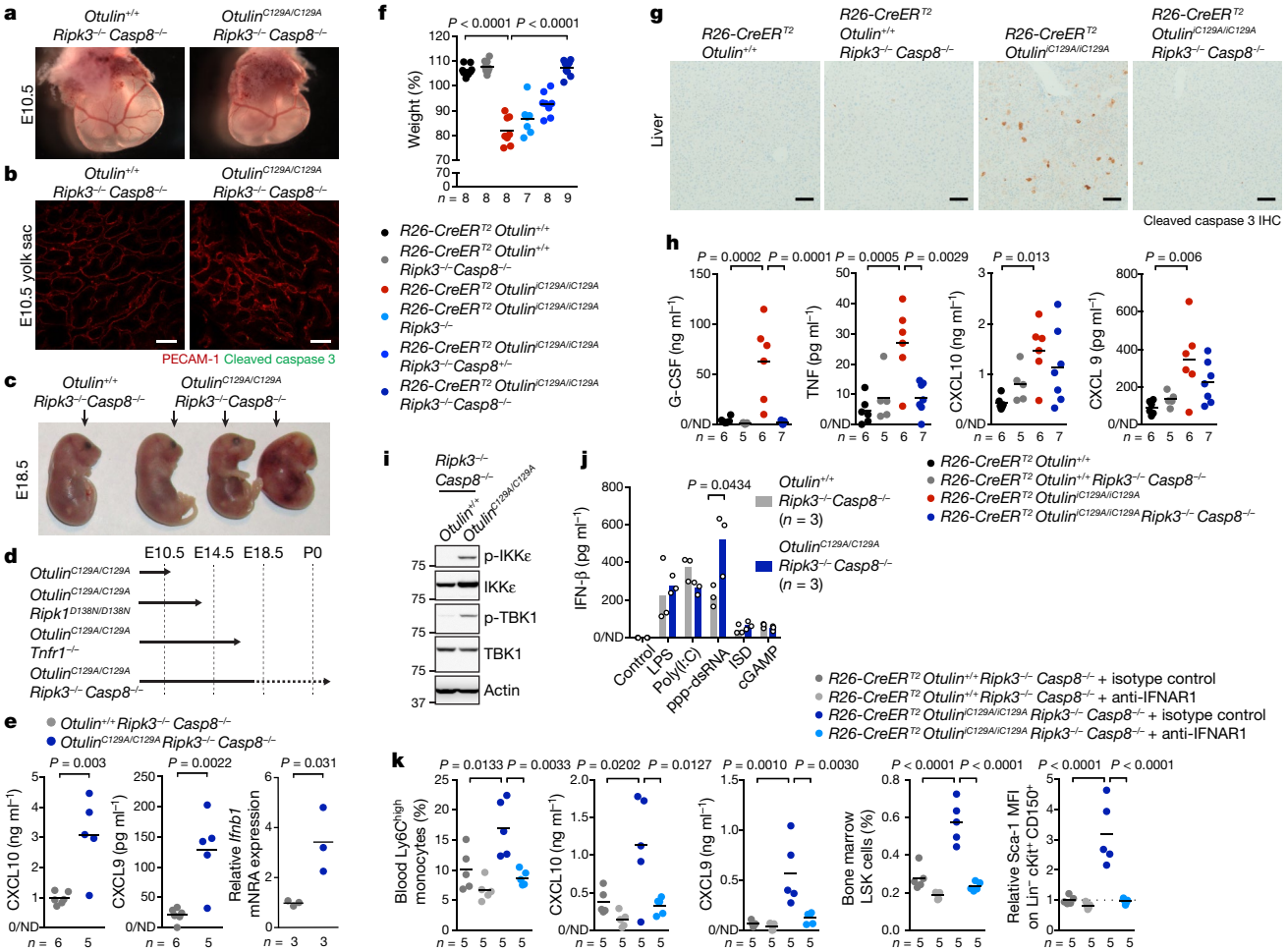


Fig. 3 | OTULIN suppresses cell death and production of type I IFN. **a**, E10.5 embryos ($n = 4$ *Otulin*$^{+/+}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$, $n = 7$ *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$). **b**, E10.5 yolk sacs stained for PECAM-1 (red) and cleaved caspase 3 (green) ($n = 4$ per genotype). Scale bar, 50 μm. **c**, E18.5 embryos ($n = 26$ *Otulin*$^{+/+}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$, $n = 25$ *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$). **d**, Schematic showing when different genotypes die. **e**, Serum chemokines or skin *Ifnb1* mRNA expression at E18.5. **f**–**h**, **k**, Tamoxifen-treated mice on day 5 (**f**–**h**) or 10 (**k**). Graphs indicate terminal body weight as a percentage of initial body weight (**f**), serum cytokines and chemokines (**h**, **k**) and flow cytometric analysis of whole blood and bone marrow (**k**). MFI, median fluorescent intensity. Liver sections were stained for cleaved caspase 3 (brown) (**g**) ($n = 5$ per genotype). Scale bar, 100 μm. **i**, Western blots of bone marrow-derived macrophages (BMDMs) ($n = 3$ per genotype). **j**, IFN-β secretion by BMDMs following 24 h treatment as shown on *x*-axis. Circles represent individual mice. Bars and lines represent means (**e**, **f**, **h**, **j**, **k**). *P* values, unpaired, two-tailed *t*-test (**e**, **j**) or one-way ANOVA followed by Tukey's multiple comparison test (**f**, **h**, **k**). ND, not detected.

Data Fig. 5l). Similar results have been reported for OTULIN-deficient lymphocytes[1].

Otulin$^{C129A/C129A}$ MEFs contained a small amount of phosphorylated IκBα, consistent with low-level, spontaneous activation of the NF-κB pathway (Fig. 2b, Extended Data Fig. 5m). However, NF-κB activation in response to TNF or lipopolysaccharide (LPS) was slightly muted in *Otulin*$^{C129A/C129A}$ MEFs (Fig. 2b, Extended Data Fig. 5m) and there was increased cell death (Fig. 2c, Extended Data Fig. 5n) with markers of both necroptosis (phospho-RIPK3 Thr231 and Ser232 and phospho-MLKL Ser345) and apoptosis (cleaved caspase 3) (Extended Data Fig. 5o). The RIPK1 inhibitor Nec-1 and/or the pan-caspase inhibitor Z-VAD-FMK (zVAD) reduced but did not completely block TNF-induced cell death (Fig. 2c). Notably, *Otulin*$^{C129A/C129A}$ MEFs were killed in a TNFR1-dependent manner by zVAD alone, with RIPK3 autophosphorylation and phospho-MLKL Ser345 indicating necroptosis (Extended Data Fig. 5n, o). Therefore, *Otulin*$^{C129A/C129A}$ MEFs were probably killed by autocrine TNF.

Collectively, these results imply that OTULIN inactivation destabilizes the TNFR1-associated complex I required for NF-κB activation and instead promotes the formation of cytosolic cell death signalling complexes. Indeed, TNF-induced complex I in *Otulin*$^{C129A/C129A}$ MEFs contained less HOIP and SHARPIN, less linear polyubiquitin, and less ubiquitinated RIPK1 than in *Otulin*$^{+/+}$ MEFs (Fig. 2d). LUBAC was not detected after complex I was treated with USP2, consistent with impaired recruitment of LUBAC into complex I rather than a failure to detect ubiquitinated LUBAC (Extended Data Fig. 6a). These alterations coincided with enhanced formation of a cytosolic complex comprised of FADD, caspase 8, TRADD, phospho-RIPK1 Ser166 and Thr169, and phospho-RIPK3 Thr231 and Ser232 (Fig. 2e). This cytosolic complex was also more abundant in *Otulin*$^{C129A/C129A}$ MEFs treated with zVAD or TNF and zVAD. Similar results have been reported for cells deficient in LUBAC components[4,14,15], suggesting that the main role of OTULIN is to maintain LUBAC function by suppressing its auto-ubiquitination.

To determine whether aberrant cell death caused embryonic lethality in *Otulin*$^{C129A/C129A}$ mice, as in HOIP-deficient mice[4], we genetically ablated RIPK1 enzymatic activity (*Ripk1*$^{D138N/D138N}$), TNFR1 (*Tnfr1*$^{-/-}$), or RIPK3 (*Ripk3*$^{-/-}$) and caspase 8 (*Casp8*$^{-/-}$). Expression of catalytically inactive RIPK1 and loss of TNFR1 delayed embryonic lethality from around E10.5 to around E14.5 and E16.5, respectively (Extended Data Fig. 6b–e). Therefore, activation of RIPK1 by TNFR1 contributes to, but is not entirely responsible for, the death of *Otulin*$^{C129A/C129A}$ embryos. Neither RIPK3 deficiency nor *Casp8* heterozygosity delayed lethality to E14.5 (Extended Data Fig. 6f–i), but the combined loss of caspase 8 and RIPK3 sustained *Otulin*$^{C129A/C129A}$

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

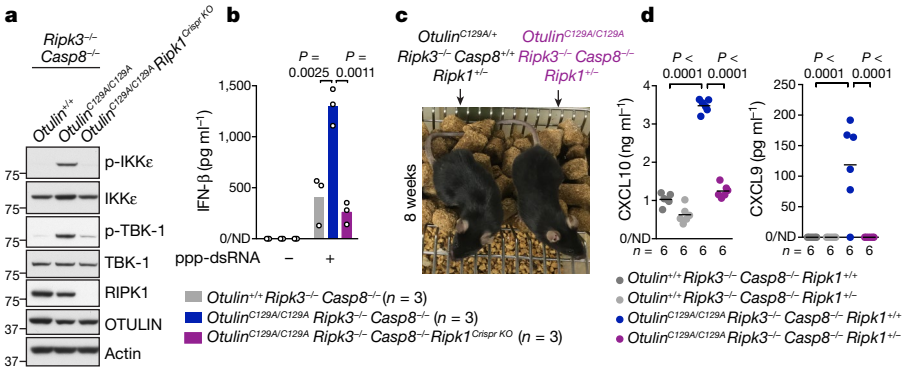


Fig. 4 | RIPK1-dependent perinatal lethality and type I IFN induction in *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ mice. a, Western blots of BMDMs. Results are representative of four independent experiments. **b**, IFN-β secretion by BMDMs following 24 h treatment with ppp-dsRNA. **c**, Littermates aged 8 weeks ($n = 5$ *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$*Ripk1*$^{+/-}$ mice (age 6–13 weeks) and control littermates). **d**, Serum chemokines at E18.5. Circles represent individual mice. Bars and lines represent means (**b**, **d**). *P* values, one-way ANOVA followed by Tukey's multiple comparison test (**b**, **d**).

embryos to the perinatal period despite them having increased linear polyubiquitin (Fig. 3a–c, Extended Data Fig. 6j–m). These data imply that aberrant cell death rather than defective Wnt signalling[3] triggers lethality in *Otulin*$^{C129A/C129A}$ embryos before E18.5 (Fig. 3d). OTULIN activity was also not required for the expression of Wnt target genes in intestinal organoid cultures (Extended Data Fig. 6n).

Some *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ embryos were oedematous at E18.5 (Fig. 3c), and newborn pups (P0) were smaller than their littermates and often found dead (Extended Data Fig. 6m). Perinatal lethality has also been reported for *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ mice with hypomorphic LUBAC due to SHARPIN mutation[16]. *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ mice had no major histological lesions at E18.5 or P0 (Extended Data Fig. 7a–c), but serum from these mice at E18.5 contained elevated amounts of the interferon (IFN)-inducible chemokines CXCL10 and CXCL9 (Fig. 3e, Extended Data Fig. 7d). *Ifnb1* mRNA was elevated in skin from E18.5 *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ mice (Fig. 3e), so we speculated that OTULIN might also suppress pathways producing type I IFN.

Aberrant cell death and TNF signalling also drove auto-inflammation in *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ mice after tamoxifen treatment. TNF blockade or TNFR1 deficiency largely suppressed the previously observed weight loss, elevated serum cytokines and chemokines, loss of haematopoietic progenitor cells, and thymic atrophy in these mice. Small intestine and liver injury was also less severe (Extended Data Fig. 7e–m). The combined loss of RIPK3 and caspase 8 was even better at ameliorating these phenotypes (Fig. 3f–h, Extended Data Fig. 8) but, notably, led to elevation of Ly6C^{high} blood monocytes (Extended Data Fig. 9a). In addition, although most pro-inflammatory cytokines, including G-CSF, TNF and IL-6, reverted to baseline levels in these mice (Fig. 3h, Extended Data Fig. 8b), the chemokines CXCL10 and CXCL9 remained elevated (Fig. 3h) and IFN-β trended upwards (Extended Data Fig. 8b). Therefore, auto-inflammation triggered by OTULIN inactivation stems largely from excessive cell death, but as seen during mouse development, blocking this death unveils an unexpected role for OTULIN in suppressing the production of chemokines linked to type I IFN.

Consistent with OTULIN modulating IFN induction, *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ BMDMs showed spontaneous phosphorylation and activation of IKKε and TBK-1, which are the kinases required for type I IFN production[17,18] (Fig. 3i). The cells also produced more IFN-β than control BMDMs after stimulation with the RIG-I ligand 5′ triphosphate double-stranded RNA (ppp-dsRNA) (Fig. 3j). CRISPR–Cas9-mediated deletion of *Hoip* in *Otulin*$^{+/+}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ BMDMs elicited a similar phenotype (Extended Data Fig. 9b, c), consistent with OTULIN and LUBAC[19,20] functioning in a linear pathway to suppress the induction of type I IFN when intracellular dsRNA is sensed. Accordingly, when tamoxifen-treated *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ and control mice were examined on day 10 rather than day 5, IFN-β, CXCL10, CXCL9, and Ly6C^{high} monocytes remained elevated (Extended Data Fig. 9d, e). Moreover, bone marrow LSK (Lin^{-} Sca-1^{+}c-Kit^{+}) cells were increased and expressed more Sca-1, which are known effects of type I IFNs[21,22] (Extended Data Fig. 9f). Indeed, an IFNAR1 (IFN-α/β receptor 1)-blocking antibody reduced Ly6C^{high} monocytes, bone marrow LSK cells, and chemokine levels (Fig. 3k, Extended Data Fig. 9g). Five weeks after tamoxifen treatment, *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ mice had myeloid-driven splenomegaly, liver and gut inflammation, and more IFN-induced chemokines in their blood than *R26-CreER*^{T2} *Otulin*$^{+/+}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ mice (Extended Data Fig. 10a–c). Collectively, our results indicate that OTULIN enzymatic activity is critical for suppressing both cell death and type I IFN production.

RIPK1 has been implicated in the type I IFN response to intracellular dsRNAs[23–25], so we investigated whether IFN signalling after OTULIN inactivation required RIPK1. CRISPR–Cas9-mediated deletion of *Ripk1* in *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ BMDMs reduced spontaneous activation of IKKε and TBK-1, and normalized IFN production induced by ppp-dsRNA (Fig. 4a, b). RIPK1 also contributed to the perinatal lethality of *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ mice because some *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ *Ripk1*$^{+/-}$ mice survived for up to 3 months, even though they were smaller than control littermates (Fig. 4c, Extended Data Fig. 10d, e). Notably, *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ *Ripk1*$^{-/-}$ mice were not found and their cause of death remains unknown. This result was unexpected because *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ *Ripk1*$^{-/-}$ mice are viable[26,27]. Nonetheless, E18.5 *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ *Ripk1*$^{+/-}$ embryos contained less CXCL9 and CXCL10 in their blood than *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ embryos (Fig. 4d, Extended Data Fig. 10f). These data indicate that the combined absence of targeted linear ubiquitination, caspase 8 and RIPK3 causes detrimental RIPK1-dependent type I IFN induction. Therefore, OTULIN emerges as a key regulator of pathways that are critical for antiviral defence by modulating cell death and type I IFN production (Extended Data Fig. 10g).

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0256-2.

Received: 30 January 2018; Accepted: 15 May 2018;
Published online: 27 June 2018

1. Damgaard, R. B. et al. The deubiquitinase OTULIN is an essential negative regulator of inflammation and autoimmunity. *Cell* **166**, 1215–1230.e20 (2016).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

2. Zhou, Q. et al. Biallelic hypomorphic mutations in a linear deubiquitinase define otulipenia, an early-onset autoinflammatory disease. *Proc. Natl Acad. Sci. USA* **113**, 10127–10132 (2016).
3. Rivkin, E. et al. The linear ubiquitin-specific deubiquitinase gumby regulates angiogenesis. *Nature* **498**, 318–324 (2013).
4. Peltzer, N. et al. HOIP deficiency causes embryonic lethality by aberrant TNFR1-mediated endothelial cell death. *Cell Reports* **9**, 153–165 (2014).
5. Keusekotten, K. et al. OTULIN antagonizes LUBAC signaling by specifically hydrolyzing Met1-linked polyubiquitin. *Cell* **153**, 1312–1326 (2013).
6. Fiil, B. K. et al. OTULIN restricts Met1-linked ubiquitination to control innate immune signaling. *Mol. Cell* **50**, 818–830 (2013).
7. Schaeffer, V. et al. Binding of OTULIN to the PUB domain of HOIP controls NF-κB signaling. *Mol. Cell* **54**, 349–361 (2014).
8. Hrdinka, M. et al. CYLD limits Lys63- and Met1-Linked ubiquitin at receptor complexes to regulate innate immune signaling. *Cell Reports* **14**, 2846–2858 (2016).
9. Elliott, P. R. et al. Molecular basis and regulation of OTULIN–LUBAC interaction. *Mol. Cell* **54**, 335–348 (2014).
10. Draber, P. et al. LUBAC-recruited CYLD and A20 regulate gene activation and cell death by exerting opposing effects on linear ubiquitin in signaling complexes. *Cell Reports* **13**, 2258–2272 (2015).
11. Elliott, P. R. & Komander, D. Regulation of Met1-linked polyubiquitin signalling by the deubiquitinase OTULIN. *FEBS J.* **283**, 39–53 (2016).
12. Alva, J. A. et al. VE-Cadherin-Cre-recombinase transgenic mouse: a tool for lineage analysis and gene deletion in endothelial cells. *Dev. Dyn.* **235**, 759–767 (2006).
13. Seibler, J. et al. Rapid generation of inducible mouse mutants. *Nucleic Acids Res.* **31**, e12 (2003).
14. Haas, T. L. et al. Recruitment of the linear ubiquitin chain assembly complex stabilizes the TNF-R1 signaling complex and is required for TNF-mediated gene induction. *Mol. Cell* **36**, 831–844 (2009).
15. Gerlach, B. et al. Linear ubiquitination prevents inflammation and regulates immune signalling. *Nature* **471**, 591–596 (2011).
16. Rickard, J. A. et al. TNFR1-dependent cell death drives inflammation in Sharpin-deficient mice. *eLife* **3**, e03464 (2014).
17. Fitzgerald, K. A. et al. IKKepsilon and TBK1 are essential components of the IRF3 signaling pathway. *Nat. Immunol.* **4**, 491–496 (2003).
18. Sharma, S. et al. Triggering the interferon antiviral response through an IKK-related pathway. *Science* **300**, 1148–1151 (2003).
19. Inn, K.-S. et al. Linear ubiquitin assembly complex negatively regulates RIG-I- and TRIM25-mediated type I interferon induction. *Mol. Cell* **41**, 354–365 (2011).
20. Belgnaoui, S. M. et al. Linear ubiquitination of NEMO negatively regulates the interferon antiviral response through disruption of the MAVS-TRAF3 complex. *Cell Host Microbe* **12**, 211–222 (2012).
21. Essers, M. A. G. et al. IFNα activates dormant haematopoietic stem cells *in vivo*. *Nature* **458**, 904–908 (2009).
22. Sato, T. et al. Interferon regulatory factor-2 protects quiescent hematopoietic stem cells from type I interferon-dependent exhaustion. *Nat. Med.* **15**, 696–700 (2009).
23. Balachandran, S., Thomas, E. & Barber, G. N. A FADD-dependent innate immune mechanism in mammalian cells. *Nature* **432**, 401–405 (2004).
24. Rajput, A. et al. RIG-I RNA helicase activation of IRF3 transcription factor is negatively regulated by caspase-8-mediated cleavage of the RIP1 protein. *Immunity* **34**, 340–351 (2011).
25. Michallet, M.-C. et al. TRADD protein is an essential component of the RIG-like helicase antiviral pathway. *Immunity* **28**, 651–661 (2008).
26. Rickard, J. A. et al. RIPK1 regulates RIPK3-MLKL-driven systemic inflammation and emergency hematopoiesis. *Cell* **157**, 1175–1188 (2014).
27. Dillon, C. P. et al. RIPK1 blocks early postnatal lethality mediated by caspase-8 and RIPK3. *Cell* **157**, 1189–1202 (2014).

Acknowledgements We thank B. Hough, E. Magee, M. Garcia and K. Cherry for animal husbandry; I. Peng and the Genentech genetic analysis, histology, necropsy, immunohistochemistry, clinical pathology and FACS laboratories for technical assistance; and J. Lai and S. Stawicki for managing antibody generation.

Reviewer information *Nature* thanks C. Griffin and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions K.H. designed and performed experiments with assistance from K.E.W., A.N., A.Mu., D.L.D., A.Ma., and F.d.S.M.; M.R.-G., R.J.N. and S.W. generated the *Otulin*$^{C129A/+}$ and *Hoip*$^{IC879S/+}$ mice; Y.Z., E.V., and D.S.K. performed mass spectrometry; J.Z. and W.P.L. performed in vivo antibody blocking experiments; J.H. performed quantitative image analysis; Y.-C.H. generated the phospho-RIPK1 Ser166, Thr169 antibody; L.G.K. performed confocal microscopy; J.D.W. analysed histological data; D.V., K.N. and V.M.D. contributed to experimental design; K.H. and K.N. wrote the manuscript with input from all authors.

Competing interests All authors were employees of Genentech.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0256-2.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0256-2.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to K.N., V.M.D.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Mice. The Genentech institutional animal care and use committee responsible for ethical compliance approved all animal protocols. *Ripk3*$^{-/-}$ (ref. [28]), *Casp8*$^{-/-}$ (ref. [29]), *Ripk1*$^{D138N/D138N}$ (ref. [29]), *Tnfr1*$^{-/-}$ (ref. [30]), *Ripk1*$^{-/-}$ (ref. [29]), *Cdh5-Cre* (ref. [12]) and *R26-CreER*^{T2} (ref. [13]) mice have been described previously. Taconic generated the *Otulin*$^{iC129A/+}$ mice from C57BL/6 NTac embryonic stem (ES) cells. In brief, a loxP-flanked cassette containing an FRT-Neo-FRT selection marker, wild-type *Otulin* cDNA (exons 4–7) and 3′ UTR followed by a human growth hormone polyadenylation signal and an F3-Puro-F3 selection marker was inserted into the *Otulin* locus 5′ to a mutated exon 4 encoding the C129A exchange (TGT to GCA). The Neo and Puro cassettes were excised by breeding to a Flp-deleter strain (Taconic).

Otulin$^{C129A/+}$ and *Hoip*$^{iC879S/+}$ mice were generated at Genentech using C57BL/6N ES cells. In brief, to generate *Otulin*$^{C129A/+}$ mice, TGT was changed to GCT in exon 4. An FRT-flanked Neo cassette placed upstream of exon 4 was excised in ES cells using Flp, leaving a single FRT 121 bp upstream of the modified exon 4. The *Hoip*$^{iC879S/+}$ allele was obtained by inserting a cassette containing a loxP, wild-type *Hoip* cDNA (exons 16–21), a human growth hormone 3′ UTR followed by a 4× polyadenylation signal, an FRT-Neo-FRT selection marker and a second loxP, into the *Hoip* locus 57 bp 5′ to a mutated exon 16 encoding the C879S (TGC to AGC) exchange. The Neo cassette was excised in ES cells using Flp. All alleles were maintained on a C57BL/6N genetic background.

Otulin^{iC129A} genotyping primers (5′-TGG GAA GGG ACT CTG A-3′ and 5′-TTT GCT ACT GTT GTA AAC TGT A-3′) amplified 176-bp wild-type, 284-bp iC129A and 256-bp C129A DNA fragments. *Otulin*^{C129A} genotyping primers (5′-CAC TGT TGC AGA TGC TGA GAT TTC-3′ and 5′-CAC TTC CTC ATA GCC CTT GG-3′) amplified 227-bp wild-type and 261-bp C129A DNA fragments. *Hoip*^{iC879S} genotyping primers (5′-CTG CTG AGT GGT CTG AT-3′, 5′-CAC TAT GGA CCT CCT TCT G-3′ and 5′-CGA GGC AGA ACA AGT CT-3′) amplified 360-bp wild-type, 457-bp iC879S and 394-bp C879S DNA fragments.

The *R26-CreER*^{T2} allele was kept in a heterozygous state. CreER^{T2} nuclear translocation in vivo was achieved by intraperitoneal (IP) injections of 3 × 40 mg/kg (Fig. 1e, f, 3f–h, Extended Data Figs. 2–4, 7e–m, 8, 9a), 3 × 80 mg/kg (Fig. 3k, Extended Data Fig. 9d–g) or 5 × 80 mg/kg (Extended Data Fig. 10a–c) tamoxifen (Sigma-Aldrich) dissolved in sunflower seed oil (Sigma-Aldrich). Dosing was once daily on consecutive days. Animals were analysed 5 days (Fig. 1e, f, 3f–h, Extended Data Figs. 2–4, 7e–m, 8, 9, 10 days (Fig. 3k, Extended Data Fig. 9d–g) or 35 days (Extended Data Fig. 10a–c) after the first tamoxifen injection.

TNF blockade was achieved by daily IP injections with either 500 μg TNFR2-Fc protein (Genentech) or 500 μg IgG2a anti-ragweed isotype control antibody (Genentech) starting one day before the first tamoxifen injection. IFNAR1 blockade was achieved by daily subcutaneous injections of either 10 mg/kg IFNAR1 blocking antibody (Leinco Technologies) or 10 mg/kg IgG1 anti-Gp120 isotype control antibody (Genentech) starting one day before the first injection of tamoxifen. Adult mice for tamoxifen treatment studies ranged in age from 6 to 18 weeks.

To generate radiation-induced bone marrow chimaeras, CD45.1 B6.SJL-Ptprca Pepcb/BoyJ (Jackson Laboratory) or CD45.2 *R26-CreER*^{T2} *Otulin*$^{+/+}$ and *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ mice were lethally irradiated and reconstituted by intravenous injection of 5 × 10^6 bone marrow cells from CD45.2 *R26-CreER*^{T2} *Otulin*$^{+/+}$ and *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ or CD45.1 B6.SJL-Ptprca Pepcb/BoyJ mice (Jackson Laboratory), respectively. Reconstitution was confirmed by flow cytometric analysis of peripheral blood 6–8 weeks after irradiation, and then mice were dosed with 3 × 40 mg/kg IP injections of tamoxifen.

Serum alanine aminotransferase (ALT), aspartate aminotransferase (AST) and total bilirubin were measured in a serum chemistry analyser (Beckman Coulter AU480) and haematology assays were run on a Sysmex XT-2000iV automated haematology analyser (cell numbers and proportions for neutrophils, monocytes, eosinophils and basophils were combined and shown as myeloid cells).

For timed pregnancies, embryos were designated E0.5 the morning a vaginal plug was detected. Animals that were euthanized because of fight wounds, malocclusion or dystocia were excluded from survival curves (Extended Data Fig. 10e).

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Immunohistochemistry, immunofluorescence and TUNEL staining. Formalin-fixed paraffin-embedded tissue sections were labelled with rabbit anti-cleaved caspase 3 (Asp175) (cat#9661, Cell Signaling Technologies, 0.5 μg/ml), rat anti-Gr-1 (RB6-8C5, BD Biosciences, 10 μg/ml), rat anti-F4/80 (CI:A3-1, Serotec BioRad, 10 μg/ml), or rabbit anti-phosphorylated (p)-RIPK3 Thr231, Ser232 (GEN135-35-9, Genentech, 5 μg/ml). Target antigen retrieval (Dako) and ABC Peroxidase Elite (Vector Laboratories) detection with DAB chromogen were used for Gr-1 and F4/80 immunohistochemistry (IHC). IHC for cleaved caspase 3 and p-RIPK3 was performed on the Ventana Discovery XT platform with CC1 standard antigen retrieval (Ventana). Cleaved caspase 3 was detected with the Ventana OmniMap detection system and DAB chromogen (Ventana). The p-RIPK3 signal was amplified with Ventana HQ Amplification and detected with HQ Discovery detection systems and DAB chromogen. Terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) was performed with the Trevigen TACS 2 TdT Blue Label Kit (Trevigen) on a Bond autostainer (Leica Biosystems). For quantitative image analysis, immunolabelled slides were digitally scanned with a Hamamatsu slide scanner. Quantitative image analysis was performed with the MatLab software package (MathWorks).

Whole-mount yolk sac immunofluorescence was performed as described[29]. Paraformaldehyde-fixed tissues were labelled with rat anti-PECAM-1 antibody (CD31, MEC13.3, BD Biosciences) and rabbit anti-cleaved caspase 3 antibody (cat#9661, Cell Signaling Technology) followed by Cy3-labelled donkey anti-rabbit IgG and Cy5-labelled donkey anti-rat IgG (both Jackson ImmunoResearch). Tissues were mounted with ProLong Gold anti-fade reagent with DAPI (Life Technologies) and images were acquired on a confocal microscope (Leica SPE).

Cell culture. Primary MEFs were isolated from E14.5 embryos excluding the head, intestine and liver. Dermal fibroblasts were isolated from adult tail and ear skin. Fibroblasts were grown in high-glucose Dulbecco's modified Eagle medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum, 2 mM glutaMAX (Gibco), 1× non-essential amino acids solution (Gibco), 100 U/ml penicillin, and 100 μg/ml streptomycin (Gibco) on tissue culture dishes pre-coated with 0.1% gelatin in PBS. CreER^{T2} nuclear translocation in fibroblasts from *R26-CreER*^{T2}, *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$, *R26-CreER*^{T2} *Hoip*$^{iC879S/iC879S}$ and *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ *Hoip*$^{iC879S/iC879S}$ mice was achieved by 72 h treatment with 1 μM 4-hydroxytamoxifen (4-OHT, Sigma-Aldrich).

Primary BMDMs and fetal liver-derived macrophages were differentiated from bone marrow and E14.5 fetal livers, respectively, for 6 days in Roswell Park Memorial Institute (RPMI) medium supplemented with 10% heat-inactivated fetal bovine serum, 2 mM glutamine, 50 μM 2-mercaptoethanol (Gibco), 100 U/ml penicillin, and 100 μg/ml streptomycin (Gibco) and 20% L929 supernatant containing M-CSF on non-treated plates. CreER^{T2} nuclear translocation in BMDMs from *R26-CreER*^{T2} *Otulin*$^{+/+}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ and *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ mice was achieved by adding 1 μM 4-OHT during differentiation.

RIPK1- or HOIP-deficient BMDMs were generated by Cas9 ribonucleoprotein electroporation. In brief, BMDMs were differentiated for 5 days in the above-mentioned DMEM medium supplemented with 50 ng/ml recombinant M-CSF (Genentech) on non-treated plates. BMDMs were harvested and extensively washed with PBS, and then 5 × 10^6 BMDMs were electroporated with recombinant Cas9 protein v2 (Invitrogen) complexed with *Ripk1*-specific guideRNAs (seed sequence/Alt-R crRNA: 5′-TAC ACG TCC GAC TTC TCC GT-3′ and 5′-AGA AGA AGG GAA CTA TTC GC-3′ annealed to Alt-R tracrRNA, IDT) or *Hoip*-specific guideRNAs (seed sequence/Alt-R crRNA: 5′-CCA CCG TGC TGC GAA AGA CG-3′ and 5′-CAT ACA ACC GTA GTA CAT CC-3′ annealed to Alt-R tracrRNA) in nucleofector solution P3 (Lonza). BMDMs were maintained in DMEM supplemented with 50 ng/ml recombinant M-CSF (Genentech) and 1 μM 4-OHT for the following 7 days.

BMDMs and MEFs were stimulated with 100 ng/ml murine TNF (Genentech), 20 μM zVAD (Promega), 30 μM Nec-1 (Tocris), 100 ng/ml (BMDMs) or 1 μg/ml (MEFs) ultra-pure LPS (Invivogen), 5 μg/ml HMW poly(I:C) (Invivogen), 1 μg/ml 5′ppp-dsRNA (Invivogen) transfected with Lipofectamine 2000 CD (Invitrogen), 2 μg/ml ISD (interferon stimulatory DNA, Invivogen) transfected with Lipofectamine 2000 CD (Invitrogen), 2 μg/ml 2′3′-cGAMP (Invivogen) and 1 μg/ml FLAG–TNF (Enzo Life Sciences). MEFs were treated with 1 μM Velcade and/or 100 nM Bafilomycin A1 (both Calbiochem).

Flt3L-dependent Hoxb8-conditionally immortalized haematopoietic progenitor cells were generated as previously described[31] by retroviral transduction of an oestrogen-regulated version of the HoxB8 oncogene into bone marrow cells followed by culturing cells in the presence of 70 ng/ml Flt3L (Peprotech) and 1 μM β-oestradiol (Sigma-Aldrich). For recombination, immortalized haematopoietic progenitor cells from *R26-CreER*^{T2} *Otulin*$^{+/+}$ and *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ mice were treated with 1 μM 4-OHT for 72 h.

Mouse organoids were established from isolated crypts collected from the entire length of the small intestine and maintained as previously described[32]. Organoids were passaged at least once and grown using IntestiCult organoid growth medium (cat#06005; StemCell Technologies). For recombination, organoids from *R26-CreER*^{T2} *Otulin*$^{+/+}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ and *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ mice were treated with 50 nM 4-OHT for 48 h. Organoids were then recovered in IntestiCult medium and expanded for 4 days.

Flow cytometry. Single-cell suspensions were prepared from organs, nonspecific antibody binding was minimized by anti-CD16/CD32 antibody (93, eBioscience), and cells were stained with fluorophore-conjugated antibodies against: CD4 (RM4-5), CD8α (53-6.7), CD11b (M1/70), CD11c (N418), CD45.1 (A20), CD45.2 (104), Ter119 (Ter119, all eBioscience), CD150 (TC15-12F12.2), Ly6C (HK1.4), Ly6G

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

(1A8), Sca-1 (D7), TCRβ (H57-597, all BioLegend), B220 (RA3-6B2), CD3ε (145-2C11), c-Kit (2B8), Gr-1 (RB6-8C5), Siglec-F (E50-2440, all BD Biosciences). Dead cells were excluded using 7-AAD (eBioscience). Samples were acquired on FACSCantoII or Fortessa machines (BD Biosciences) and analysed with FlowJo software. Doublets were excluded based on FSC-H/A and SSC-W/A characteristics and cell populations were gated as follows: LSK cells (lineage (CD11b, CD11c, Gr-1, B220, TCRβ, Ter119)$^-$ Sca-1$^+$ c-Kit$^+$), LK cells (lineage (CD11b, CD11c, Gr-1, B220, TCRβ, Ter119)$^-$ Sca-1$^-$ c-Kit$^+$), Ly6C^{high} monocytes (CD11b$^+$ Ly6G$^-$ Ly6C^{high}), neutrophils (CD11b$^+$ Ly6G$^+$) and as indicated in the Extended Data Figure legends. Gating strategies for flow cytometric analyses can be found in Supplementary Fig. 2.

MEF survival was analysed after overnight treatment by combining adherent and floating dead cells. Samples were stained with 2 μg/ml PI (Sigma-Aldrich), and acquired on FACSCantoII or Fortessa machines (BD Biosciences).

Western blotting, immunoprecipitation and mass spectrometry. For western blotting of whole cell lysates, cells were lysed or embryo tissues were mechanically disrupted in 50 mM Tris HCL pH 7.4, 150 mM NaCl, 2 mM EDTA, 0.5% (w/v) Na-Deoxycholate, 0.1% (w/v) SDS, 1% (v/v) NP40, PhosSTOP phosphatase inhibitor (Roche), complete protease inhibitor cocktail (Roche) and 1 mM PMSF (Sigma-Aldrich). For immunoprecipitation, cells were lysed in 20 mM Tris HCl pH 7.5, 135 mM NaCl, 1.5 mM $MgCl_2$, 1 mM EGTA, 1% (v/v) Triton X-100, 10% (v/v) glycerol, PhosSTOP phosphatase inhibitor, and complete protease inhibitor cocktail. 10 mM N-ethylmaleimide (NEM, Sigma-Aldrich) was added to lysis buffer if ubiquitination was analysed. Insoluble material was removed by centrifugation at 20,000*g* before addition of LDS sample buffer or IP.

Western blotting antibodies: RIPK3 (1G6.1.4, Genentech), phosphorylated RIPK3 Thr231, Ser232 (GEN135-35-9)[29], RIPK1 (10C7, Genentech), phosphorylated RIPK1 Ser166, Thr169 (GEN175-DP-A1 raised against peptide ASFKTW(pS)KL(pT)KEKD, Genentech), MLKL (3H1, Millipore), phosphorylated MLKL Ser354 (EPR9515(2), Abcam), FLAG (M2, Sigma), actin (C4, MP Biomedicals), FADD (1.28E12, Genentech), caspase 8 (1G12, Enzo Life Sciences), OTULIN (GN13.18, Genentech), HOIP (11D6H2G5, Genentech), HOIL-1 (2E2, Millipore), SHARPIN (14626-1-AP, Proteintech), linear ubiquitin (1F11/3F5/Y102L, Genentech), K48 ubiquitin (APU2.07, Genentech), LC3 (8E10, MBL) and TRADD (GN21-3, Genentech). The following antibodies were from Cell Signaling Technologies: phosphorylated ERK (cat#9101), ERK (cat#9102), phosphorylated JNK (cat#4668), JNK (cat#9258), phosphorylated IκBα (cat#2859), IκBα (cat#9242), phosphorylated p65/RelA (cat#3033), p65/RelA (cat#8242), phosphorylated p38 (cat#9211), p38 (cat#8690), RIPK1 (cat#3493), phosphorylated IKKε (cat#8766), IKKε (cat#3416), phosphorylated TBK1 (cat#5483), TBK1 (cat# 3504), caspase 3 (cat#9662), cleaved caspase 3 (cat# 9664), and cleaved caspase 8 (cat#8592).

Anti-FADD antibody (1F7, Millipore) was used for IP. Complexes were recovered with magnetic protein A/G beads (Pierce) or magnetic anti-FLAG beads (M2, Sigma-Aldrich).

To enrich for proteins modified with linear polyubiquitin, cells were lysed in 100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM EDTA, 1% NP-40, 0.5% Triton-X 100 containing complete protease inhibitor cocktail, 2 mM NEM and 150 nM FLAG-anti-linear polyubiquitin TUBEs (LifeSensors). Insoluble material was removed by centrifugation at 20,000*g* and lysates were diluted tenfold in 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM EDTA containing complete protease inhibitor cocktail, 2 mM NEM and 150 nM FLAG-anti-linear polyubiquitin TUBEs. FLAG–TUBEs were recovered with magnetic anti-FLAG beads, washed with 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM EDTA, 0.05% NP-40 containing complete protease inhibitor cocktail and 2 mM NEM, and bound proteins were eluted for 30 min with 150 μg/ml 3 × FLAG peptide (Sigma-Aldrich) by rotation at 4 °C.

For mass spectrometry-based identification of proteins modified with linear and K48-linked polyubiquitin, modified proteins were enriched sequentially with 150 nM FLAG anti-linear polyubiquitin TUBEs followed by 150 nM FLAG anti-K48-linked polyubiquitin TUBEs (LifeSensors). Samples were eluted with 3 × FLAG peptide and separated by SDS–PAGE for ~1 cm. Coomassie (Invitrogen)-stained gel lanes were each excised as a single region and subjected to in gel digestion. Gel pieces were cut into strips, with 100 μl of 50 mM ammonium bicarbonate (AMBIC) in a 50:50 mixture of acetonitrile (ACN):water, desiccated with 50 μl of 100% ACN, and then dried for 1 h in a SpeedVac. Trypsin solution (20 ng/μl in 25 mM AMBIC; Promega) was added to dried gel pieces, incubated on ice for 1 h, then incubated for 16 h at 37 °C. Digested peptides were transferred to a new tube and gel pieces extracted with 100 μl 10% ACN 0.1% trifluoroacetic acid (TFA). The peptide containing solution was dried completely in a SpeedVac and resuspended in 0.1% TFA for mass spectrometry analysis. Samples were injected onto a QExactive-HF mass spectrometer using a NanoAcquity UPLC and analysed by data dependent shotgun sequencing. MS1 scans were collected at 60,000 resolution with an AGC target of 3E6 and a maximum ion time of 60 ms and the top 10 precursor ions subjected to MS2 analysis at 15,000 resolution with an AGC target of 1E5 and max ion time of 75 ms. Raw data files were searched with semi-tryptic specificity using Mascot against the concatenated target-decoy database from UniProt (downloaded April 2015; *Mus musculus* taxonomy) and common contaminants. Oxidized methionine was permitted as a variable modification. Peptide spectral matches were filtered on a per run basis to 5% false discovery rate at the peptide level using linear discriminant analysis, and then in aggregate across all runs to a 2% false discovery rate at the protein level. Total and unique peptide counts are reported.

For IsoT, OTULIN or USP2 treatment, cells were lysed in 50 mM Tris pH 8.0, 150 mM NaCl, 1% Triton, 10% glycerol supplemented with EDTA-free protease inhibitors (Roche). IPs were washed three times with 20 mM Tris HCl pH 7.5, 135 mM NaCl, 1.5 mM $MgCl_2$, 1 mM EGTA, 1% (v/v) Triton X-100, 10% (v/v) glycerol, PhosSTOP phosphatase inhibitor, and complete protease inhibitor cocktail and then once with 50 mM Tris pH 8.0, 150 mM NaCl. Subsequently, IPs or diluted whole cell lysates were incubated with 1 μM IsoT, OTULIN or USP2 (all Boston Biochem) in assay buffer (50 mM Tris pH 8.0, 150 mM NaCl, 5 mM DTT) for 1 h at 32 °C. IsoT was pre-incubated in 50 mM Tris, pH 8.0, 150 mM NaCl and 10 mM DTT at 23 °C for 10 min, before the deubiquitination assay. IsoT treatment of 1 μg free linear tetra-ubiquitin chains (Boston Biochem) was performed under the same conditions. Reactions were quenched by adding LDS loading buffer (Invitrogen). Samples were subjected to SDS-PAGE analysis with subsequent staining using InstantBlue Protein Stain (Expedeon).

Cytokine and chemokine detection. Sera from adult mice or E18.5 embryos and cell supernatants were analysed using Premix Panel I 32-plex (Millipore) or IFN-β ELISA (PBL).

Quantitative reverse-transcription PCR. RNA from organoids, MEFs, macrophages or mechanically disrupted skin tissue was isolated using a RNeasy mini-kit (Qiagen). Quantitative reverse-transcription–PCR was performed using One-step Real-time RT–PCR mastermix (ABI). Taqman probes were from Life Technologies: *Actb* (Mm01205647_g1 or Mm02619580_g1), *Lgr5* (Mm01251801_m1), *Axin2* (Mm00443610_m1), *Ascl2* (Mm01268891_g1), *Olfm4* (Mm01320260_m1), *Sharpin* (Mm00550584_m1), *Hoil-1* (Mm00498069_m1), *Hoip* (Mm01313902_m1) and *Ifnb1* (Mm00439552_s1). Values were normalized to *Actb* transcript levels.

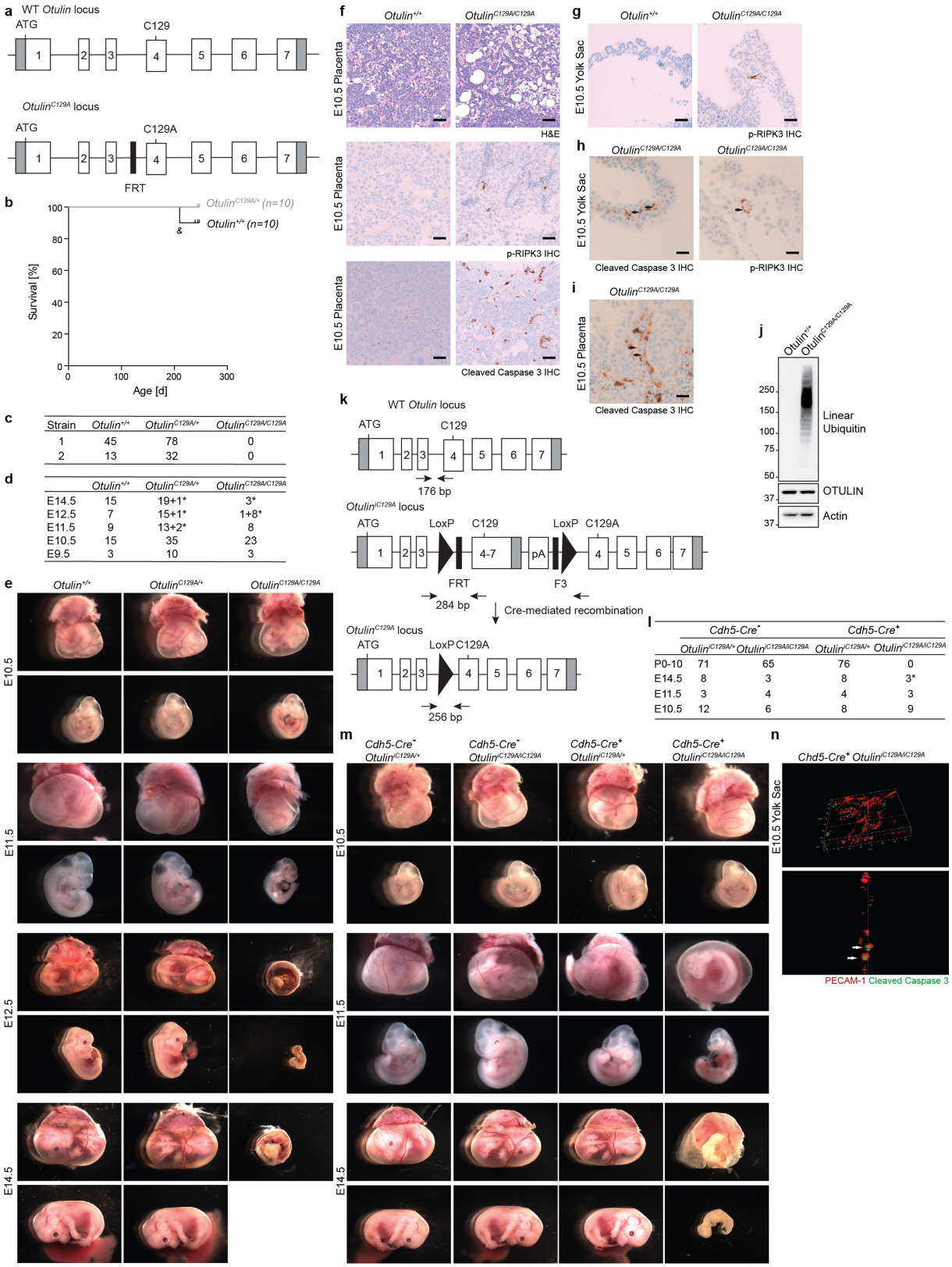
Statistics. Statistical analysis of the results was performed by unpaired, two-tailed *t*-test, by one-way ANOVA followed by Tukey's multiple comparison test or by repeated measures two-way ANOVA followed by Bonferroni's multiple comparison test. A 95% confidence interval was used for statistics and $P < 0.05$ was considered significant. All statistical analyses were performed using GraphPad Prism 6.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The data sets generated in this study are available from the corresponding authors upon reasonable request. Proteomics raw data files have been deposited to the UCSD MassIVE database (ftp://massive.ucsd.edu/MSV000081959). Gel source data can be found in Supplementary Fig. 1. Source data for all graphs are provided.

28. Newton, K., Sun, X. & Dixit, V. M. Kinase RIP3 is dispensable for normal NF-kappa Bs, signaling by the B-cell and T-cell receptors, tumor necrosis factor receptor 1, and Toll-like receptors 2 and 4. *Mol. Cell. Biol.* **24**, 1464–1469 (2004).
29. Newton, K. et al. Activity of protein kinase RIPK3 determines whether cells die by necroptosis or apoptosis. *Science* **343**, 1357–1360 (2014).
30. Peschon, J. J. et al. TNF receptor-deficient mice reveal divergent roles for p55 and p75 in several models of inflammation. *J. Immunol.* **160**, 943–952 (1998).
31. Redecke, V. et al. Hematopoietic progenitor cell lines with myeloid and lymphoid potential. *Nat. Methods* **10**, 795–803 (2013).
32. Sato, T. et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262–265 (2009).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



c

Strain	Otulin +/+	Otulin C129A/+	Otulin C129A/C129A
1	45	78	0
2	13	32	0

d

	Otulin +/+	Otulin C129A/+	Otulin C129A/C129A
E14.5	15	19+1*	3*
E12.5	7	15+1*	1+8*
E11.5	9	13+2*	8
E10.5	15	35	23
E9.5	3	10	3

l

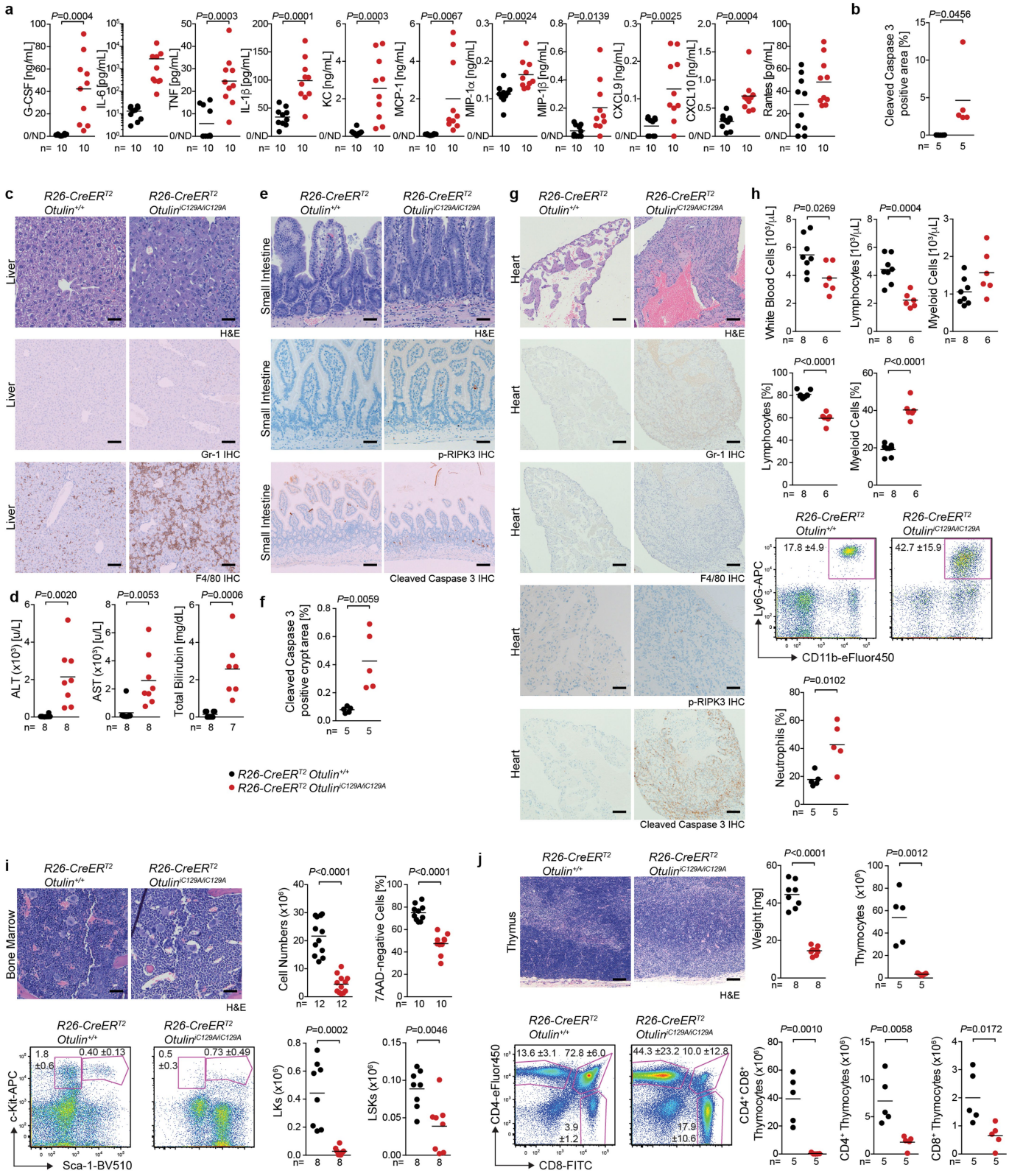
	Cdh5-Cre-		Cdh5-Cre+	
	Otulin iC129A/+	Otulin iC129A/iC129A	Otulin iC129A/+	Otulin iC129A/iC129A
P0-10	71	65	76	0
E14.5	8	3	8	3*
E11.5	3	4	4	3
E10.5	12	6	8	9

Extended Data Fig. 1 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 1 | Constitutive and conditional OTULIN inactivation using a *Cdh5-Cre* transgene causes embryonic lethality. **a**, Organization of the constitutive *Otulin*^{C129A} knock-in allele. Boxes represent exons. Untranslated regions are shaded grey. **b**, Kaplan–Meier plot of mouse survival. &, Animal euthanized owing to dermatitis with pruritus. **c**, Numbers of offspring of *Otulin*$^{C129A/+}$ parents at clipping (P0–P10). Strains were derived from independently targeted ES cell clones. **d**, Numbers of offspring from intercrossing *Otulin*$^{C129A/+}$ parents. *Embryos resorbed. **e**, Representative embryos from **d** with and without their yolk sacs. **f**–**i**, E10.5 placenta (**f**, **i**) or yolk sac (**g**, **h**) stained with haematoxylin and eosin (H&E) or labelled for phospho-RIPK3 Thr231, Ser232 or cleaved caspase 3 by immunohistochemistry (IHC) ($n = 4$ *Otulin*$^{+/+}$, $n = 3$ *Otulin*$^{C129A/C129A}$ mice in (**f**, **i**); $n = 7$ *Otulin*$^{+/+}$, $n = 3$ *Otulin*$^{C129A/C129A}$ mice in **g**, **h**). Scale bars, 100 μm (H&E staining and cleaved caspase 3 IHC in **f**), 50 μm (p-RIPK3 IHC in **f**, **g**) or 25 μm (cleaved caspase 3 IHC and p-RIPK3 IHC in **h**, **i**). Arrows indicate labelling of vessel walls (**h**, **i**). **j**, Western blots of E10.5 embryos ($n = 3$ per genotype). **k**, Organisation of the inducible *Otulin*^{iC129A} knock-in allele. Boxes represent exons. Untranslated regions are shaded grey. Arrows indicate genotyping primer locations. **l**, Offspring numbers from intercrossing *Otulin*$^{iC129A/iC129A}$ with *Cdh5-Cre*$^{+}$ *Otulin*$^{iC129A/+}$ parents. *Embryos resorbed. **m**, Representative embryos from **l** with and without their yolk sacs. **n**, E10.5 yolk sacs stained for PECAM-1 (red) and cleaved caspase 3 (green) ($n = 2$ *Cdh5-Cre*$^{+}$ *Otulin*$^{iC129A/iC129A}$). 3D projection of 17 confocal *z* stacks. The lower image shows part of an orthogonal projection. Arrows indicate cells containing cleaved caspase 3 and PECAM-1.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

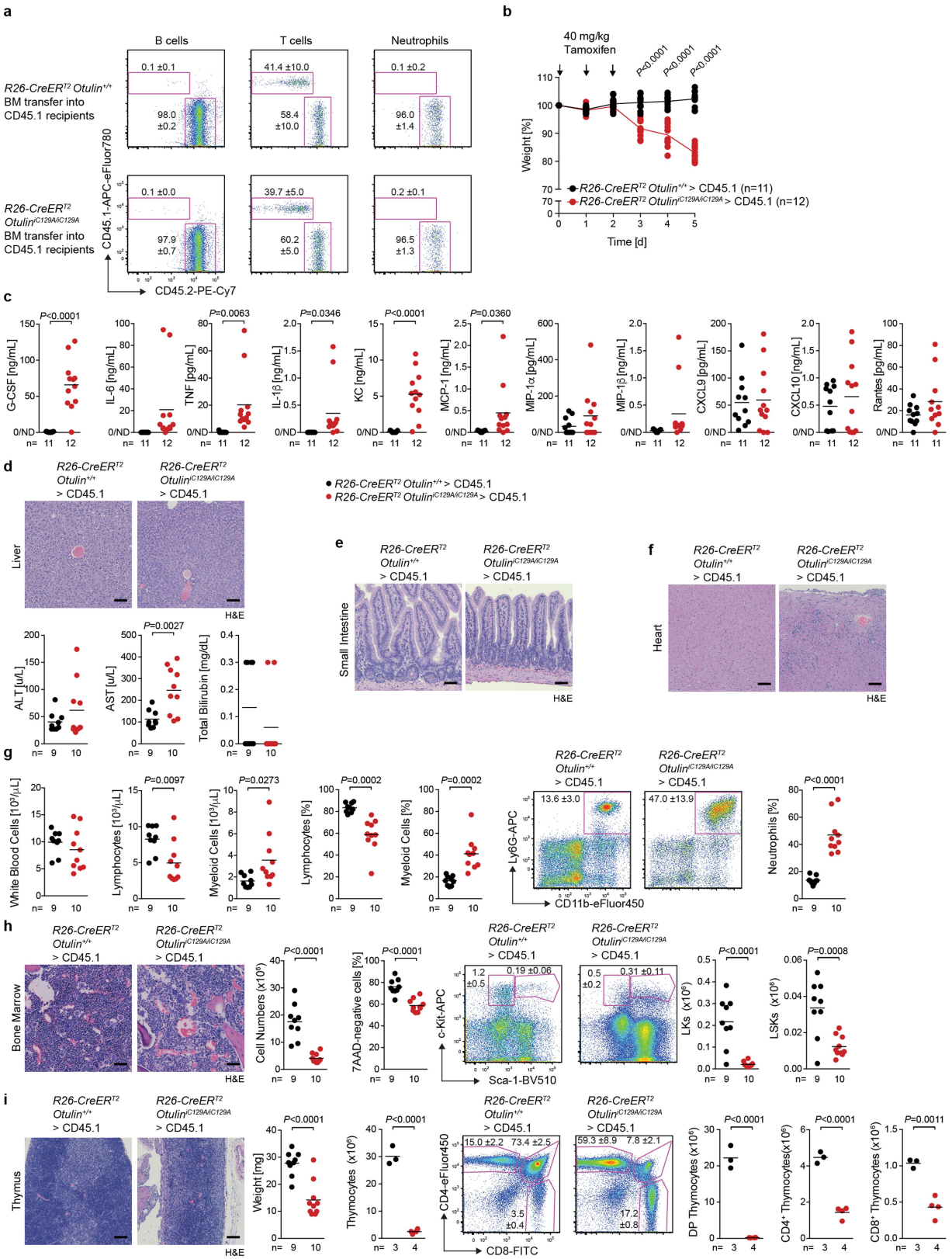


Extended Data Fig. 2 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 2 | Systemic OTULIN inactivation using the *R26-CreER^{T2}* transgene causes spontaneous inflammation and immune cell defects. Additional characterization of the day 5 tamoxifen-treated mice shown in Fig. 1e. **a**, Serum cytokine and chemokine levels. **b**, The percentage of liver tissue positive for cleaved caspase 3 by IHC. **c**, Liver stained with haematoxylin and eosin (H&E; scale bar, 50 μm; $n = 8$ per genotype) or labelled with Gr-1 or F4/80 IHC (scale bar, 100 μm; $n = 5$ per genotype). **d**, Serum ALT, AST, and total bilirubin levels. **e**, Small intestine stained with H&E (scale bar, 50 μm; $n = 8$ per genotype) or labelled by phospho-RIPK3 Thr231, Ser232 IHC (scale bar, 50 μm; $n = 5$ per genotype) or cleaved caspase 3 IHC (scale bar, 100 μm; $n = 5$ per genotype). **f**, The percentage of crypt tissue in the small intestine positive for cleaved caspase 3 by IHC. **g**, Heart tissue. Scale bar, 100 μm (H&E, $n = 8$ per genotype; and Gr-1, F4/80 and cleaved caspase 3 IHCs, $n = 5$ per genotype) or 50 μm (phospho-RIPK3 Thr231, Ser232 IHC; $n = 5$ per genotype). **h**, Blood parameters. Neutrophils ($CD11b^+$ $Ly6G^+$) are boxed in representative FACS plots. The mean ± s.d. percentage of neutrophils is indicated ($n = 5$ per genotype). **i**, Sternum stained with H&E (scale bar, 50 μm; $n = 8$ per genotype). Graphs indicate bone marrow cell numbers in two femurs and two tibia, the percentage of viable bone marrow cells not stained by 7-AAD, and numbers of bone marrow LK and LSK cells. LK and LSK cells are indicated in representative FACS plots with their mean ± s.d. frequency ($n = 8$ per genotype). **j**, Thymus stained with H&E (scale bar, 100 μm; $n = 8$ per genotype). Graphs indicate thymus weights and total thymocyte numbers. $CD4^+CD8^+$, $CD4^+$ and $CD8^+$ thymocytes are indicated in representative FACS plots with their mean ± s.d. frequency ($n = 5$ per genotype). Circles represent individual mice. Lines represent means; *P* values, unpaired, two-tailed *t*-test (**a**, **b**, **d**, **f**, **h**, **i**, **j**).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

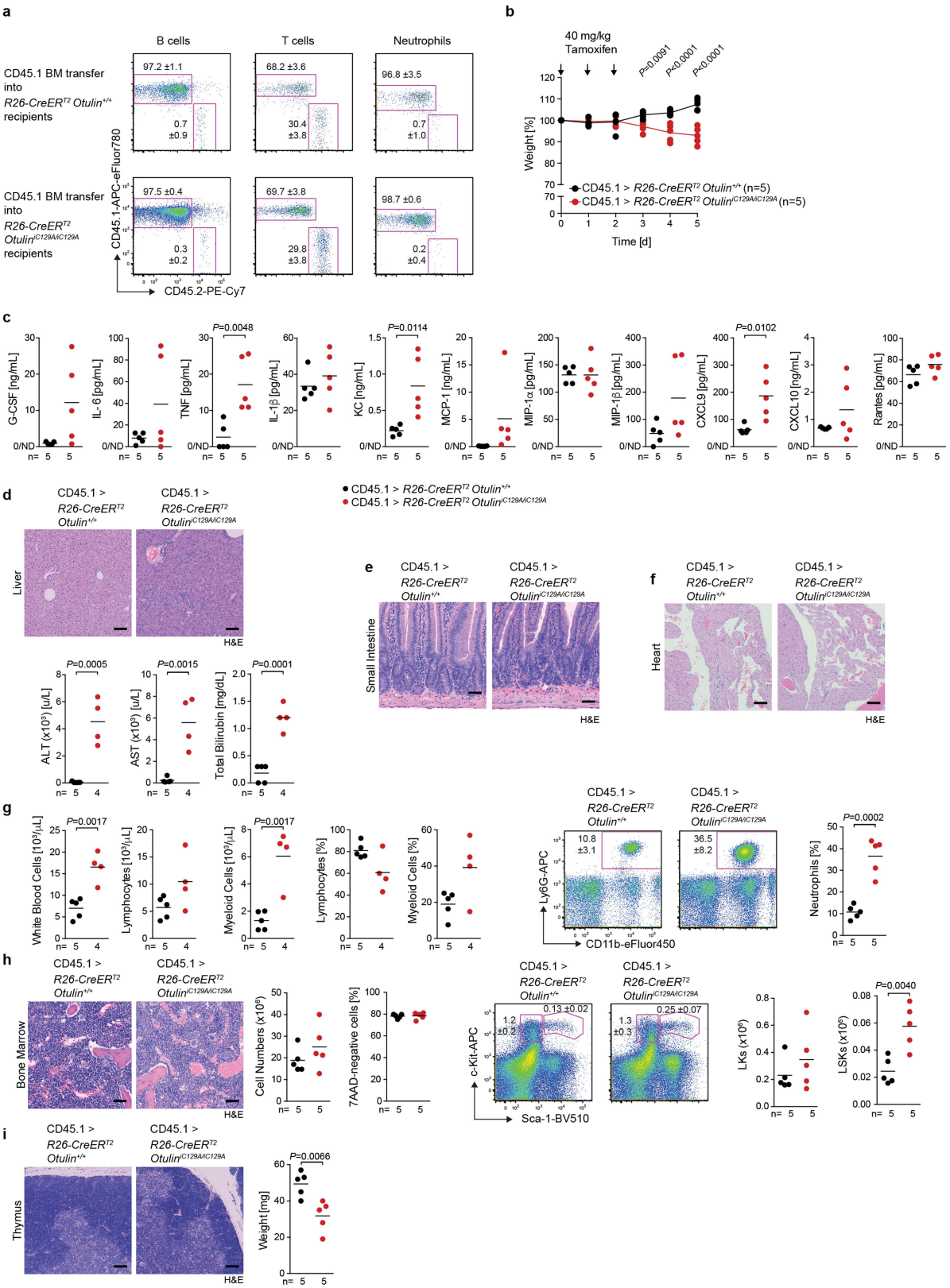


Extended Data Fig. 3 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 3 | OTULIN inactivation in the haematopoietic compartment causes spontaneous inflammation. a, FACS plots indicate the efficiency with which irradiated CD45.1$^+$ mice were reconstituted with bone marrow (BM) from CD45.2$^+$ *R26-CreER^{T2} Otulin$^{+/+}$* or *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$* mice. B cells (B220$^+$ CD3ε$^-$), T cells (B220$^-$ CD3ε$^+$) and neutrophils (CD11b$^+$ Ly6G$^+$) in whole blood were examined. Numbers represent the mean percentage ± s.d. ($n = 11$ *R26-CreER^{T2} Otulin$^{+/+}$*, $n = 12$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$*). **b**, Body weights of tamoxifen-treated mice plotted as a percentage of initial body weight. Lines represent the mean. *P* values, repeated measures two-way ANOVA followed by Bonferroni's multiple comparison test. **c–i**, Characterization of the mice in **b** on day 5. **c**, Serum cytokines and chemokines. **d**, Liver (scale bar, 100 μm; $n = 9$ *R26-CreER^{T2} Otulin$^{+/+}$*, $n = 10$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$*) and serum ALT, AST, and total bilirubin. **e**, Small intestine (scale bar, 50 μm; $n = 9$ *R26-CreER^{T2} Otulin$^{+/+}$*, $n = 10$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$*). **f**, Heart (scale bar, 100 μm; $n = 9$ *R26-CreER^{T2} Otulin$^{+/+}$*, $n = 10$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$*). **g**, Blood parameters. Numbers in representative FACS plots indicate the mean ± s.d. neutrophil frequency ($n = 9$ *R26-CreER^{T2} Otulin$^{+/+}$*, $n = 10$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$*). **h**, Sternum (scale bar, 50 μm; $n = 9$ *R26-CreER^{T2} Otulin$^{+/+}$*, $n = 10$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$*). Graphs indicate bone marrow cell numbers in two femurs and two tibia, the percentage of viable bone marrow cells not stained by 7-AAD, and numbers of bone marrow LK and LSK cells. LK and LSK cells are indicated in representative FACS plots with their mean ± s.d. frequency ($n = 9$ *R26-CreER^{T2} Otulin$^{+/+}$*, $n = 10$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$*). **i**, Thymus (scale bar, 100 μm; $n = 6$ *R26-CreER^{T2} Otulin$^{+/+}$*, $n = 6$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$*). Graphs indicate thymus weights and total thymocyte numbers. CD4$^+$CD8$^+$, CD4$^+$ and CD8$^+$ thymocytes are indicated in representative FACS plots with their mean ± s.d. frequency ($n = 3$ *R26-CreER^{T2} Otulin$^{+/+}$*, $n = 4$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$*). Circles represent individual mice. Lines represent means; *P* values, unpaired, two-tailed *t*-test (**c**, **d**, **g–i**).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

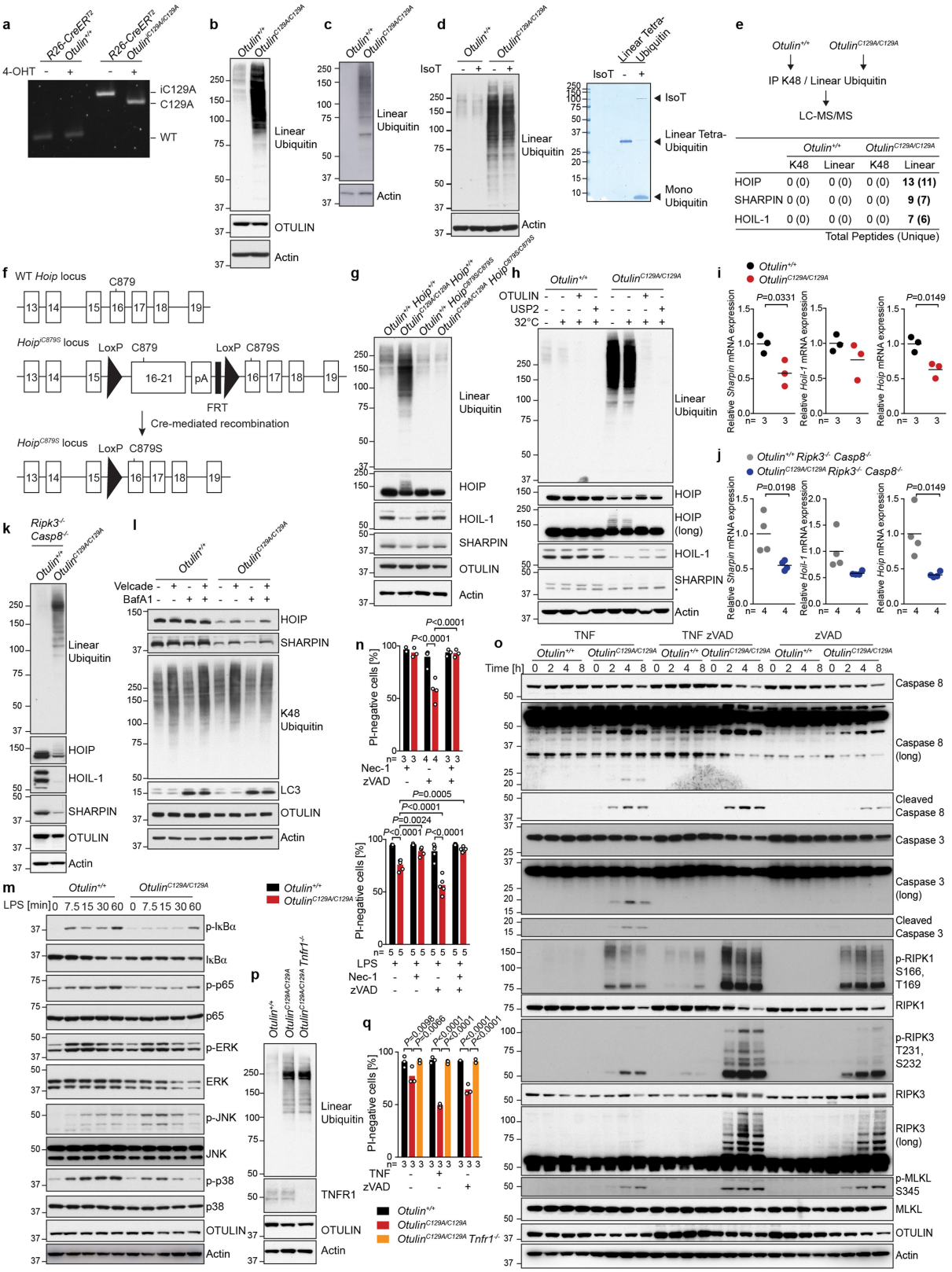


Extended Data Fig. 4 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 4 | OTULIN inactivation outside the haematopoietic compartment causes mild inflammation. a, FACS plots indicate the efficiency with which irradiated CD45.2^{+} *R26-CreER^{T2} Otulin$^{+/+}$* and *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$* mice were reconstituted with BM from CD45.1^{+} mice. B cells (B220^{+} TCRβ^{-}), T cells (B220^{-} TCRβ^{+}) and neutrophils (CD11b^{+} Ly6G^{+}) in whole blood were examined. Numbers represent the mean percentage ± s.d. ($n = 12$ per genotype). **b**, Body weights of tamoxifen-treated mice plotted as a percentage of initial body weight. Lines represent means; *P* values, repeated measures two-way ANOVA followed by Bonferroni's multiple comparison test. **c**–**i**, Characterization of the mice in **b** on day 5. **c**, Serum cytokines and chemokines. **d**, Liver (scale bar, 100 μm; $n = 5$ per genotype) and serum ALT, AST and total bilirubin. **e**, Small intestine (scale bar, 50 μm; $n = 5$ per genotype). **f**, Heart (scale bar, 100 μm; $n = 5$ per genotype). **g**, Blood parameters. Numbers in representative FACS plots indicate the mean neutrophil frequency ± s.d. ($n = 5$ per genotype). **h**, Sternum (scale bar, 50 μm; $n = 5$ per genotype). Graphs indicate bone marrow cell numbers in two femurs and two tibia, the percentage of viable bone marrow cells not stained by 7-AAD, and numbers of bone marrow LK and LSK cells. LK and LSK cells are indicated in representative FACS plots with their mean ± s.d. frequency ($n = 5$ per genotype). **i**, Thymus (scale bar, 100 μm; $n = 5$ per genotype). Graph indicates thymus weights. Circles represent individual mice. Lines represent means; *P* values, unpaired, two-tailed *t*-test (**c**, **d**, **g**–**i**).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

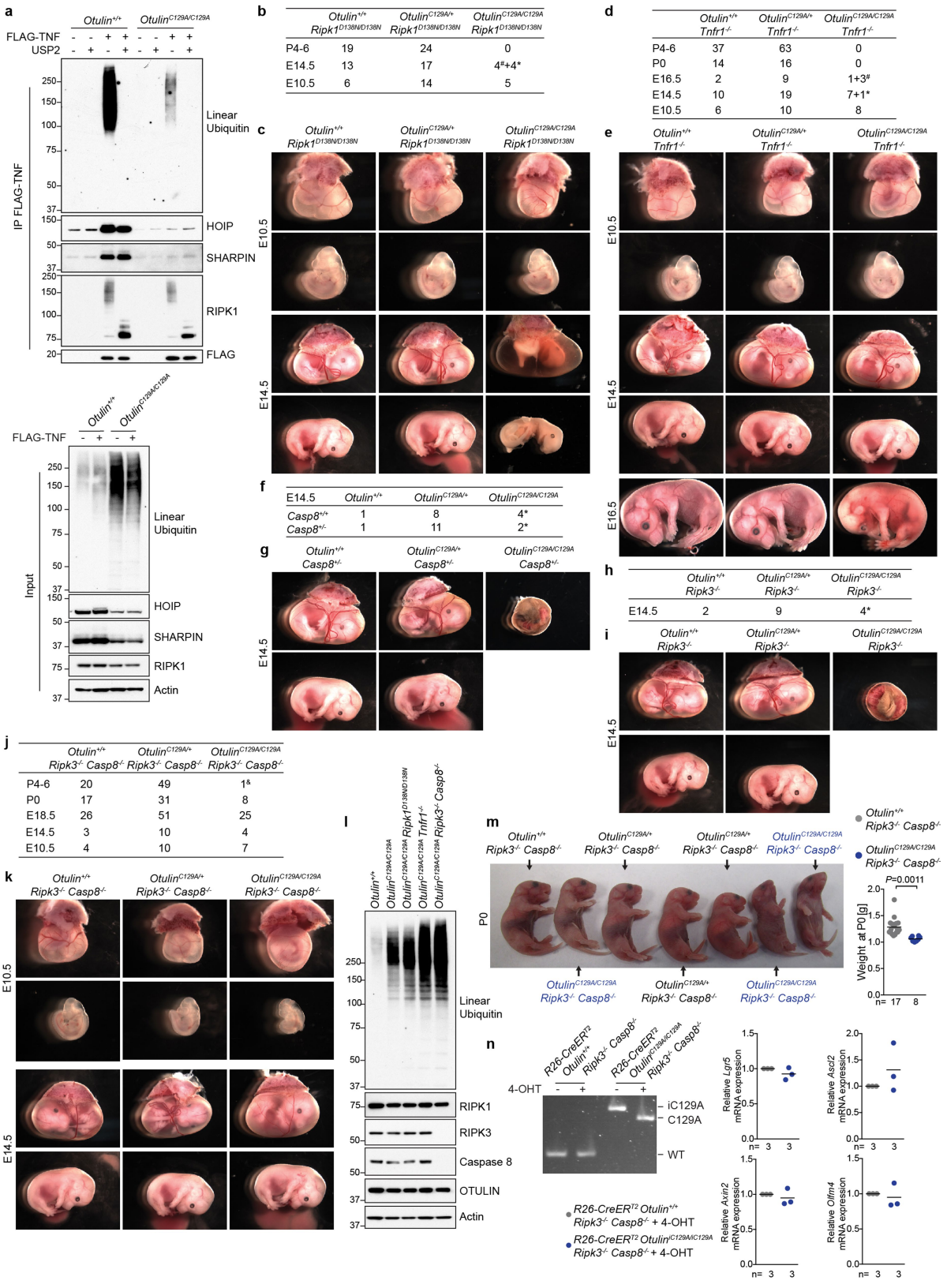


Extended Data Fig. 5 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 5 | OTULIN inactivation leads to LUBAC auto-ubiquitination, a reduction in LUBAC components, muted NF-κB activation and enhanced cell death upon stimulation. a, PCR on genomic DNA isolated from primary MEFs treated or not treated with 4-OHT. **b**, **c**, **g**, **k**, **m**, **o**, **p**, Western blots of MEFs (**b**, **m**, **o**, **p**), immortalized haematopoietic progenitors (**c**), dermal fibroblasts (**g**) and fetal liver-derived macrophages (**k**). **d**, Western blots of MEF lysates treated or not treated with IsoT (left). Coomassie-stained gel of free linear tetra-ubiquitin chains treated or not treated with IsoT (right). **e**, Total (unique) peptide counts by mass spectrometry after affinity purification of linear or K48-linked ubiquitin from haematopoietic progenitors. **f**, Organization of the conditional *Hoip*^{iC879S} knock-in allele. Boxes represent exons. **h**, Western blots of MEF lysates with or without OTULIN or USP2 treatment and with or without incubation at 32 °C. *Nonspecific band. **i**, **j**, Graphs of relative *Sharpin, Hoil-1* and *Hoip* mRNA expression in MEFs (**i**) or fetal liver-derived macrophages (**j**). Each circle represents cells from a different embryo. Lines represent means; *P* values, unpaired, two-tailed *t*-test (**i**, **j**). **l**, Western blots of MEFs treated for 4 h as indicated. **n**, **q**, Graphs indicate the percentage of MEFs that were viable and not stained by PI after overnight treatment. Each circle represents cells from a different embryo. Bars represent means; *P* values, one-way ANOVA followed by Tukey's multiple comparison test (**n**, **q**). Results are representative of two independent experiments (**d**, **h**, **l**, **m**, **o**) or cells from three mice per genotype (**a**–**c**, **g**, **k**, **p**).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

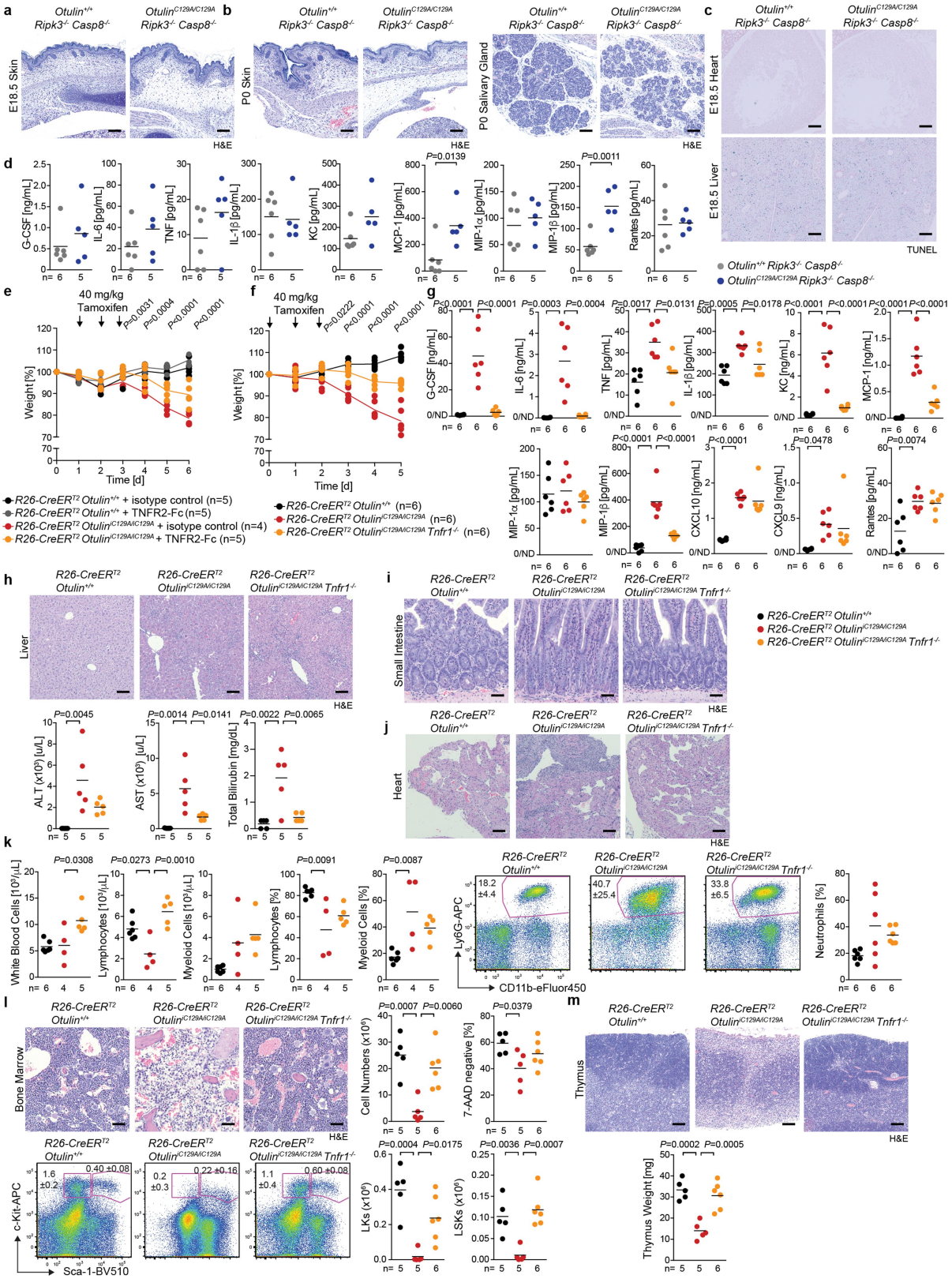


Extended Data Fig. 6 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 6 | Cell death mediated by TNFR1, RIPK1 catalytic activity, caspase 8 and RIPK3 contributes to the embryonic lethality caused by OTULIN inactivation. a, Western blots of MEFs and their immunoprecipitates (IP) with or without USP2 treatment. Results are representative of two independent experiments. **b**, Offspring numbers from intercrossing *Otulin*$^{C129A/+}$ *Ripk1*$^{D138N/D138N}$ parents. *Embryos resorbed; $^{\#}$embryos grossly abnormal. **c**, Representative embryos from **b** with and without their yolk sacs. **d**, Offspring numbers from intercrossing *Otulin*$^{C129A/+}$ *Tnfr1*$^{-/-}$ parents. *Embryos resorbed; $^{\#}$Embryos grossly abnormal. **e**, Representative embryos from **d** with and without their yolk sacs. **f**, E14.5 embryo numbers from intercrossing *Otulin*$^{C129A/+}$ *Casp8*$^{+/-}$ with *Otulin*$^{C129A/+}$ *Casp8*$^{+/+}$ parents. *Embryos resorbed. **g**, Representative embryos from **f** with and without their yolk sacs. **h**, E14.5 embryo numbers from intercrossing *Otulin*$^{C129A/+}$ *Ripk3*$^{-/-}$ parents. *Embryos resorbed. **i**, Representative embryos from **h** with and without their yolk sacs. **j**, Offspring numbers from intercrossing *Otulin*$^{C129A/+}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ parents. $^{\&}$Not found at wean. **k**, Representative embryos from **j** with and without their yolk sacs. **l**, Western blots of E10.5 embryos ($n = 1$ embryo per genotype). **m**, A representative P0 litter from **j**. Graph indicates newborn body weights. Circles represent individual pups. Lines represent means. *P* value, unpaired, two-tailed *t*-test. **n**, PCR on genomic DNA isolated from primary intestinal organoids treated with 4-OHT or untreated ($n = 3$ per genotype). Graphs indicate relative *Lgr5*, *Ascl2*, *Axin2* and *Olfm4* mRNA levels in intestinal organoids treated with 4-OHT. Circles represent cells from different mice. Lines represent means.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

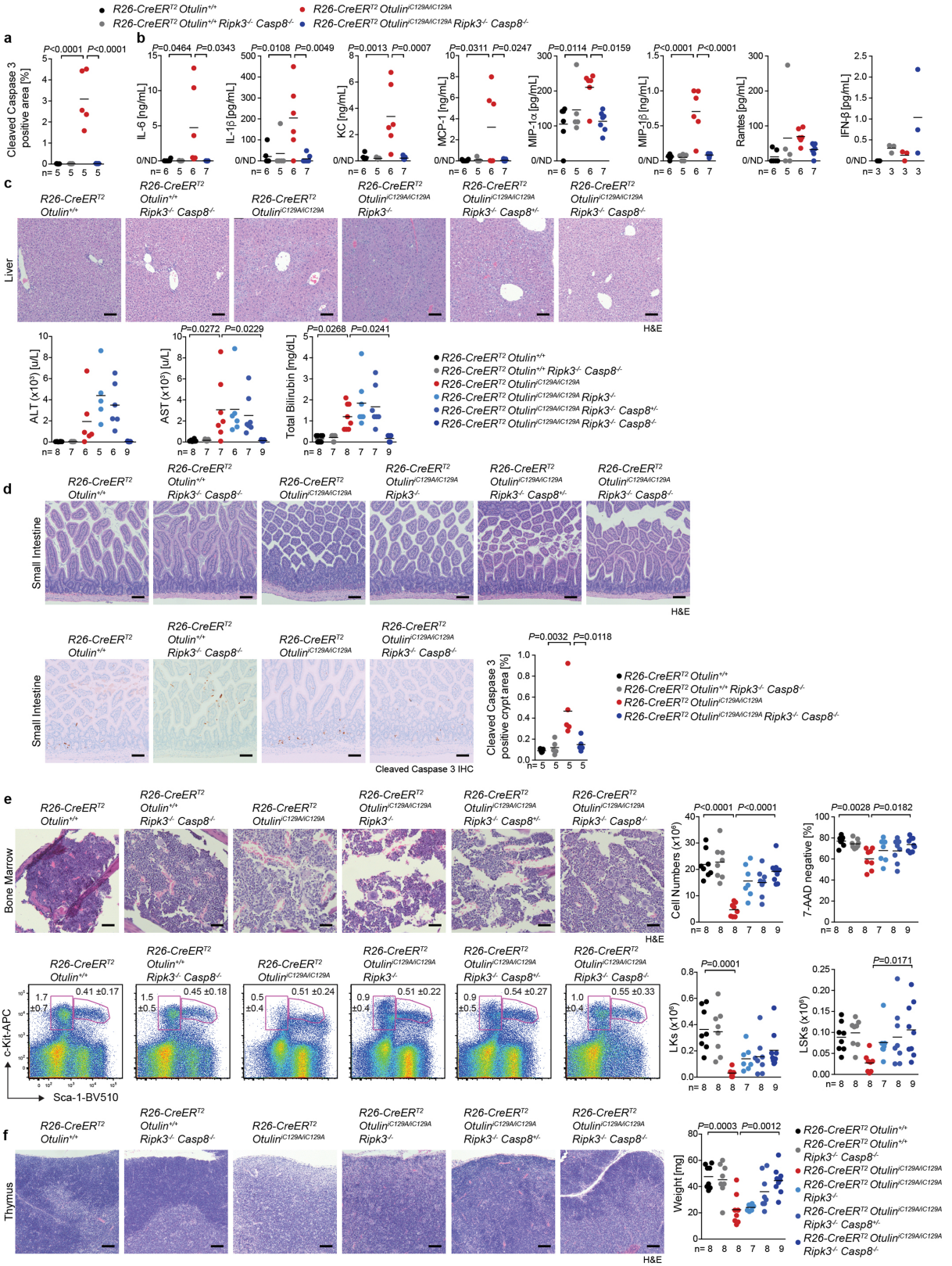


Extended Data Fig. 7 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 7 | OTULIN inactivation causes perinatal lethality in the absence of RIPK3 and caspase 8, and TNF signalling-dependent inflammation in adult mice. a, **b**, E18.5 and P0 skin or salivary gland (scale bar, 100 μm; E18.5, $n = 8$ *Otulin*$^{+/+}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$, $n = 5$ *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$; P0, $n = 3$ *Otulin*$^{+/+}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$, $n = 3$ *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$). Oedema was noted in three out of five *Otulin*$^{C129A/C129A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ mice at E18.5 and in two out of three mice at P0. **c**, E18.5 heart and liver stained for nicked DNA by TUNEL assay (scale bar, 200 μm; $n = 3$ per genotype). **d**, E18.5 serum cytokine and chemokine levels. Each circle represents one embryo. **e**, Body weights of tamoxifen-treated mice plotted as a percentage of initial body weight. Lines represent means. *P* values, repeated measures two-way ANOVA followed by Bonferroni's multiple comparison test. *P* values comparing *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ isotype control and TNFR2-Fc treated animals are shown. **f**, Body weights of tamoxifen-treated mice plotted as a percentage of initial body weight. Lines represent the mean. *P* values, repeated measures two-way ANOVA followed by Bonferroni's multiple comparison test. *P* values comparing *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ and *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ *Tnfr1*$^{-/-}$ mice are shown. **g**–**m**, Characterization of the mice in **f** on day 5. **g**, Serum cytokine and chemokine levels. **h**, Liver (scale bar, 100 μm; $n = 5$ *R26-CreER*^{T2} *Otulin*$^{+/+}$, $n = 5$ *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$, $n = 6$ *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ *Tnfr1*$^{-/-}$) and serum ALT, AST, and total bilirubin. **i**, Small intestine (scale bar, 50 μm; $n = 5$ *R26-CreER*^{T2} *Otulin*$^{+/+}$, $n = 5$ *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$, $n = 6$ *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ *Tnfr1*$^{-/-}$). **j**, Heart (scale bar, 100 μm; $n = 5$ *R26-CreER*^{T2} *Otulin*$^{+/+}$, $n = 5$ *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$, $n = 6$ *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ *Tnfr1*$^{-/-}$). **k**, Blood parameters. Numbers in representative FACS plots indicate the mean neutrophil frequency ± s.d. ($n = 6$ per genotype). **l**, Sternum (scale bar, 50 μm; $n = 5$ *R26-CreER*^{T2} *Otulin*$^{+/+}$, $n = 5$ *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$, $n = 6$ *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ *Tnfr1*$^{-/-}$). Graphs indicate bone marrow cell numbers in two femurs and two tibia, the percentage of viable bone marrow cells not stained by 7-AAD, and numbers of bone marrow LK and LSK cells. LK and LSK cells are indicated in representative FACS plots with their mean ± s.d. frequency ($n = 5$ *R26-CreER*^{T2} *Otulin*$^{+/+}$, $n = 5$ *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$, $n = 6$ *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ *Tnfr1*$^{-/-}$). **m**, Thymus (scale bar, 100 μm; $n = 5$ *R26-CreER*^{T2} *Otulin*$^{+/+}$, $n = 5$ *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$, $n = 6$ *R26-CreER*^{T2} *Otulin*$^{iC129A/iC129A}$ *Tnfr1*$^{-/-}$). Graphs indicate thymus weights. Circles represent individual mice. Lines represent means; *P* values, unpaired, two-tailed *t*-test (**d**) or one-way ANOVA followed by Tukey's multiple comparison test (**g**, **h**, **k**–**m**).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

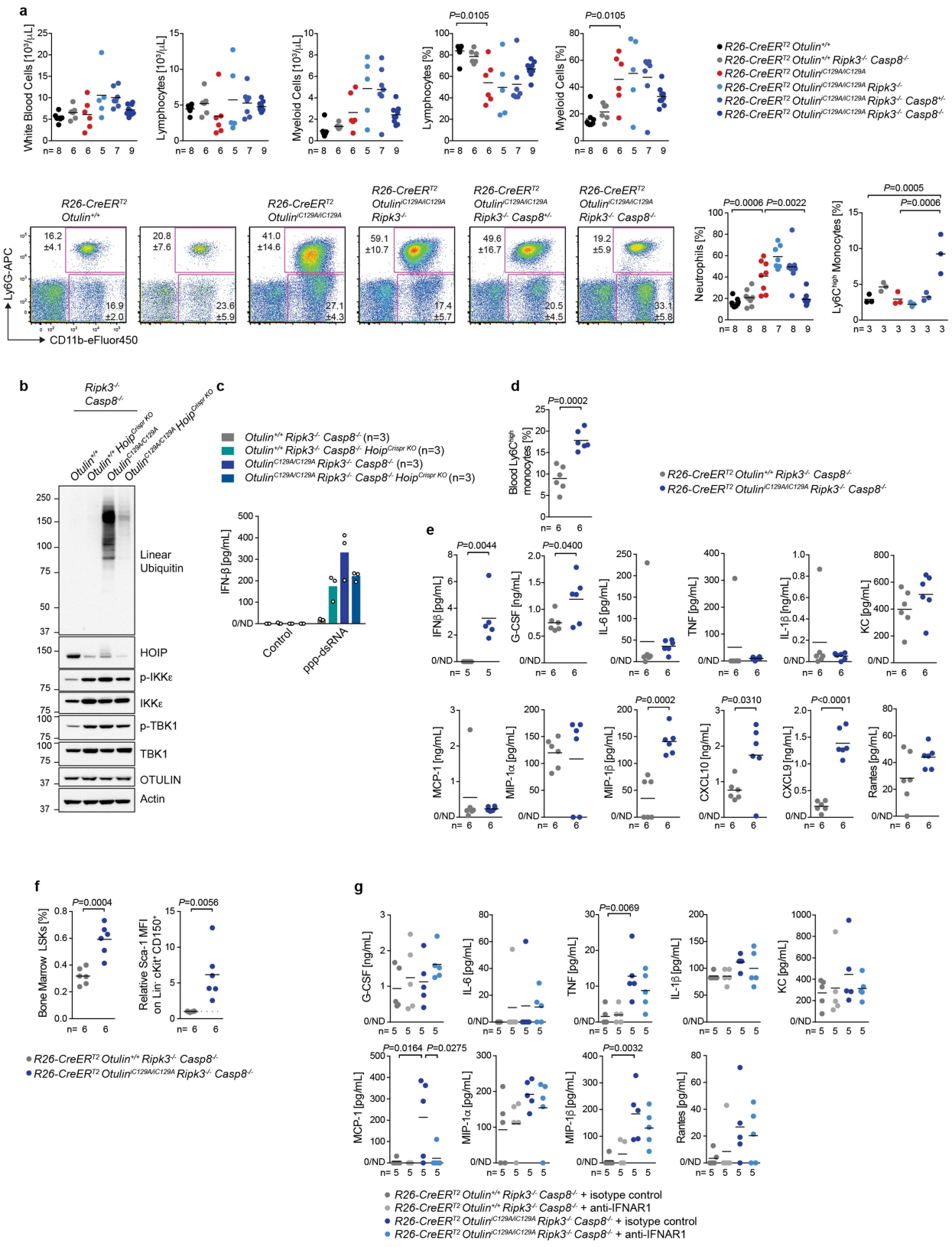


Extended Data Fig. 8 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 8 | RIPK3 and caspase 8 deficiency rescues spontaneous inflammation and most immune cell defects caused by OTULIN inactivation in the adult. Additional characterization of the day 5 tamoxifen-treated mice shown in Fig. 3f. **a**, Graph of the percentage of liver tissue that was positive for cleaved caspase 3 by IHC ($n = 5$ per genotype). **b**, Serum cytokines and chemokines. **c**, Liver (scale bar, 100 μm; $n = 8$ *R26-CreER^{T2} Otulin$^{+/+}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{+/+}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$*, $n = 7$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$ Casp8$^{+/-}$*, $n = 9$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*) and serum ALT, AST, and total bilirubin. **d**, Small intestine (scale bar, 100 μm; $n = 8$ *R26-CreER^{T2} Otulin$^{+/+}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{+/+}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$*, $n = 7$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$ Casp8$^{+/-}$*, $n = 9$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*). Lower panels indicate labelling of cleaved caspase 3 (brown, $n = 5$ per genotype) and the graph shows the percentage of crypt tissue that was positive. **e**, Sternum (scale bar, 50 μm; $n = 8$ *R26-CreER^{T2} Otulin$^{+/+}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{+/+}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$*, $n = 7$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$ Casp8$^{+/-}$*, $n = 9$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*). Graphs indicate bone marrow cell numbers in two femurs and two tibia, the percentage of viable bone marrow cells not stained by 7-AAD, and numbers of bone marrow LK and LSK cells. LK and LSK cells are indicated in representative FACS plots with their mean ± s.d. frequency ($n = 8$ *R26-CreER^{T2} Otulin$^{+/+}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{+/+}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$*, $n = 7$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$ Casp8$^{+/-}$*, $n = 9$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*). **f**, Thymus (scale bar, 100 μm; $n = 8$ *R26-CreER^{T2} Otulin$^{+/+}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{+/+}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$*, $n = 7$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$ Casp8$^{+/-}$*, $n = 9$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*). Graph indicates thymus weights. Circles represent individual mice. Lines represent means; *P* values, one-way ANOVA followed by Tukey's multiple comparison test (**a**–**f**).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

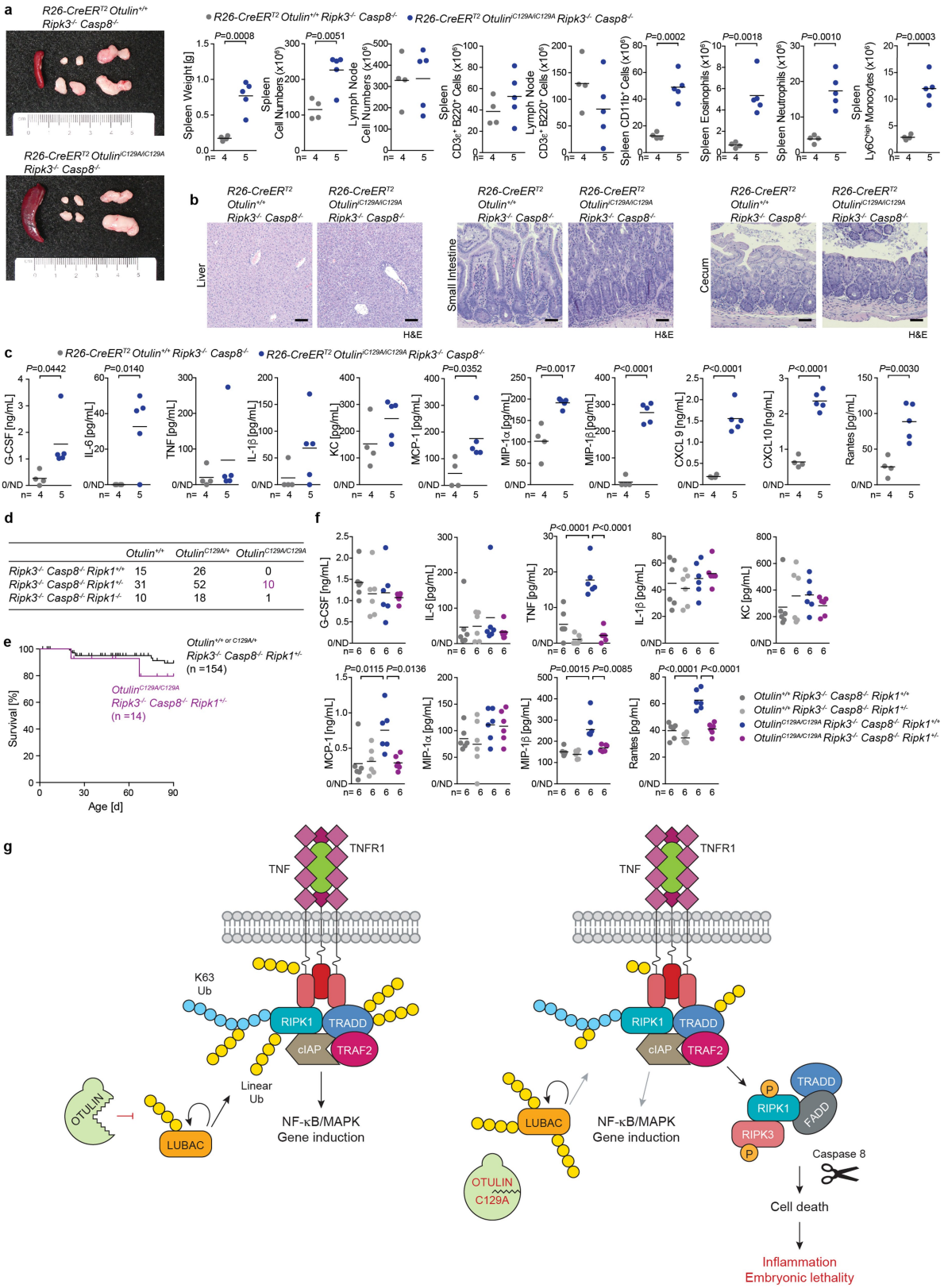


Extended Data Fig. 9 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 9 | OTULIN inactivation in adults lacking RIPK3 and caspase 8 elicits a type I IFN signature. Immune cell phenotypes of the day 5 tamoxifen-treated mice shown in Fig. 3f. **a**, Blood parameters. Numbers in representative FACS plots indicate the mean ± s.d. neutrophil frequency ($n = 8$ *R26-CreER^{T2} Otulin$^{+/+}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{+/+}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$*, $n = 7$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$*, $n = 8$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$ Casp8$^{+/-}$*, $n = 9$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*). **b**, Western blots of BMDMs. Results are representative of cells from three mice of each genotype. **c**, IFN-β secretion by BMDMs following 24 h treatment as indicated. Circles represent cells from individual mice. Bars represent means. **d**–**g**, Characterization of day 10 tamoxifen-treated mice. **d**, Graph indicates CD11b^{+} Ly6G^{-} Ly6C^{high} blood cells. **e**, Serum cytokines and chemokines. **f**, Graphs indicate the percentage of LSK bone marrow cells and SCA-1 expression (MFI, median fluorescent intensity) on the indicated bone marrow subsets. **g**, Serum cytokines and chemokines in the mice shown in Fig. 3k. Circles represent individual mice and lines represent means (**a**, **d**–**g**). *P* values, unpaired, two-tailed *t*-test (**d**–**f**) or one-way ANOVA followed by Tukey's multiple comparison test (**a**, **g**).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 10 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 10 | OTULIN inactivation in the combined absence of RIPK3 and caspase 8 causes RIPK1-dependent inflammation. **a**–**c**, Characterization of day 35 tamoxifen-treated mice. **a**, Representative spleen and lymph nodes (inguinal, brachial/axillary, mandibular/cervical and mesenteric lymph; $n = 4$ *R26-CreER^{T2} Otulin$^{+/+}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*, $n = 5$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*). Graphs indicate spleen weight, total spleen or lymph node cellularity, and numbers of different cellular subsets (CD3ε^{+} B220^{+} T cells, CD11b^{+} myeloid cells, CD11b^{+} Ly6G^{-} SiglecF^{+} SSC-A^{high} eosinophils, CD11b^{+} Ly6G^{+} neutrophils and CD11b^{+} Ly6G^{-} SiglecF^{-} SSC-A^{low} Ly6C^{high} monocytes). **b**, Liver (scale bar, 100 μm), small intestine (scale bar, 50 μm) and caecum (scale bar, 50 μm) ($n = 4$ *R26-CreER^{T2} Otulin$^{+/+}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*, $n = 5$ *R26-CreER^{T2} Otulin$^{iC129A/iC129A}$ Ripk3$^{-/-}$ Casp8$^{-/-}$*). **c**, Serum cytokines and chemokines. **d**, Offspring numbers at P4–6 from intercrossing *Otulin$^{C129A/+}$ Ripk3$^{-/-}$ Casp8$^{-/-}$ Ripk1$^{+/-}$* parents. **e**, Kaplan–Meier plot of mouse survival. Note that some mice were obtained independent from intercrossing *Otulin$^{C129A/+}$ Ripk3$^{-/-}$ Casp8$^{-/-}$ Ripk1$^{+/-}$* parents and are not included in **d**. **f**, E18.5 serum cytokines and chemokines. Circles represent individual mice (**a**, **c**) or individual embryos (**f**). Lines represent means (**a**, **c**, **f**). *P* values, unpaired, two-tailed *t*-test (**a**, **c**) or one-way ANOVA followed by Tukey's multiple comparison test (**f**). **g**, Model of how OTULIN limits cell death, embryonic lethality and inflammation. OTULIN counters auto-ubiquitination of the E3 ligase LUBAC. Stimulation of TNFR1 by its cognate ligand TNF initiates assembly of complex I containing TRADD, TRAF2, cIAP1/2 and RIPK1. Ubiquitin chains generated by the E3 ligases cIAP1 and cIAP2 recruit LUBAC but not OTULIN into complex I, leading to the modification of complex I components with linear polyubiquitin and activation of NF-κB and MAPK signalling for gene induction. OTULIN inactivation results in LUBAC auto-ubiquitination, a reduction in total LUBAC, and reduced LUBAC recruitment into complex I. These changes promote formation of cytosolic cell death signalling complexes containing RIPK1, RIPK3, FADD, TRADD and caspase 8. Increased necroptotic and apoptotic cell death causes both embryonic lethality and auto-inflammation in the adult. RIPK1 mediates type I IFN induction in the combined absence of OTULIN function, caspase 8 and RIPK3.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

https://doi.org/10.1038/s41586-018-0251-7

Acquired resistance to IDH inhibition through *trans* or *cis* dimer-interface mutations

Andrew M. Intlekofer[1,2,3,4,15], Alan H. Shih[1,2,4,5,15], Bo Wang[1,2,6], Abbas Nazir[1,2], Ariën S. Rustenburg[7], Steven K. Albanese[7,8], Minal Patel[2], Christopher Famulare[2], Fabian M. Correa[1,2], Naofumi Takemoto[1,2], Vidushi Durani[1,2], Hui Liu[9], Justin Taylor[1,2,4,5], Noushin Farnoud[2,10,11], Elli Papaemmanuil[2,10,11], Justin R. Cross[9], Martin S. Tallman[4,5], Maria E. Arcila[12], Mikhail Roshal[12], Gregory A. Petsko[13], Bin Wu[14], Sung Choe[14], Zenon D. Konteatis[14], Scott A. Biller[14], John D. Chodera[7], Craig B. Thompson[6,16]*, Ross L. Levine[1,2,4,5,16]* & Eytan M. Stein[4,5,16]*

Somatic mutations in the isocitrate dehydrogenase 2 gene (*IDH2*) contribute to the pathogenesis of acute myeloid leukaemia (AML) through the production of the oncometabolite 2-hydroxyglutarate (2HG)[1–8]. Enasidenib (AG-221) is an allosteric inhibitor that binds to the IDH2 dimer interface and blocks the production of 2HG by IDH2 mutants[9,10]. In a phase I/II clinical trial, enasidenib inhibited the production of 2HG and induced clinical responses in relapsed or refractory *IDH2*-mutant AML[11]. Here we describe two patients with *IDH2*-mutant AML who had a clinical response to enasidenib followed by clinical resistance, disease progression, and a recurrent increase in circulating levels of 2HG. We show that therapeutic resistance is associated with the emergence of second-site *IDH2* mutations *in trans*, such that the resistance mutations occurred in the *IDH2* allele without the neomorphic R140Q mutation. The *in trans* mutations occurred at glutamine 316 (Q316E) and isoleucine 319 (I319M), which are at the interface where enasidenib binds to the IDH2 dimer. The expression of either of these mutant disease alleles alone did not induce the production of 2HG; however, the expression of the Q316E or I319M mutation together with the R140Q mutation *in trans* allowed 2HG production that was resistant to inhibition by enasidenib. Biochemical studies predicted that resistance to allosteric IDH inhibitors could also occur via IDH dimer-interface mutations *in cis*, which was confirmed in a patient with acquired resistance to the IDH1 inhibitor ivosidenib (AG-120). Our observations uncover a mechanism of acquired resistance to a targeted therapy and underscore the importance of 2HG production in the pathogenesis of *IDH*-mutant malignancies.

We investigated two patients with *IDH2*-mutant AML who developed acquired resistance to enasidenib, hypothesizing that we could identify molecular mechanisms of resistance to small-molecule IDH inhibition. The first patient with AML relapsed after induction with daunorubicin and cytarabine, and was refractory to subsequent decitabine treatment as evidenced by an increasing blast count and progressive neutropenia (Fig. 1a, b). The plasma 2HG concentration was increased at 3.6 μM (Fig. 1c). Next-generation sequencing of bone marrow cells aspirated demonstrated a clonal $IDH2^{R140Q}$ mutation

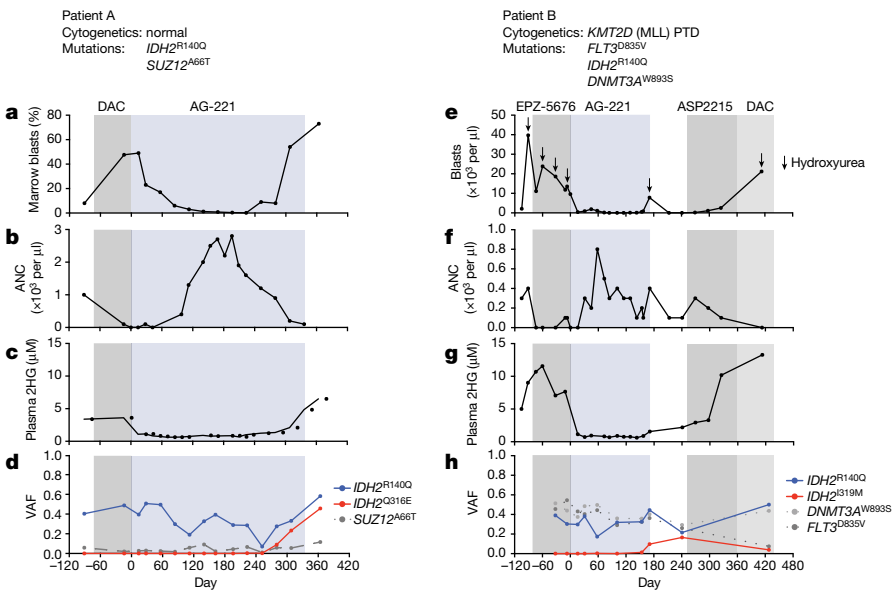


Fig. 1 | Acquired resistance to the mutant IDH2 inhibitor enasidenib (AG-221) associated with emergence of second-site mutations in *IDH2*. **a–d**, Clinical, laboratory and pathological features for patient A in relation to enasidenib (AG-221) treatment (blue box) and decitabine (DAC; grey box), including bone marrow blast percentage (**a**), blood absolute neutrophil count (ANC) (**b**), plasma 2HG concentration (**c**), and variant allele frequency (VAF) for mutations identified by targeted next-generation sequencing of bone marrow cells (**d**). **e–h**, Clinical, laboratory and pathological features for patient B in relation to enasidenib (AG-221) treatment (blue box) and other treatments (grey boxes), including blood absolute blast count (**e**), blood ANC (**f**), plasma 2HG concentration (**g**), and VAF for mutations identified by targeted next-generation sequencing of bone marrow cells (**h**). MLL, mixed lineage leukaemia; PTD, partial tandem duplication. Also see Extended Data Fig. 1.

[1]Human Oncology & Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [2]Center for Hematologic Malignancies, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [3]Lymphoma Service, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [4]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [5]Leukemia Service, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [6]Cancer Biology & Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [7]Computational & Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [8]Gerstner Sloan Kettering Graduate School, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [9]The Donald B. and Catherine C. Marron Cancer Metabolism Center, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [10]Department of Epidemiology & Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [11]Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [12]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [13]Feil Family Brain and Mind Research Institute, Weill Cornell Medical College, New York, New York, USA. [14]Agios Pharmaceuticals, Inc, Cambridge, MA, USA. [15]These authors contributed equally: Andrew M. Intlekofer, Alan H. Shih. [16]These authors jointly supervised this work: Craig B. Thompson, Ross L. Levine, Eytan M. Stein. *e-mail: thompsonc@mskcc.org; leviner@mskcc.org; steine@mskcc.org

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

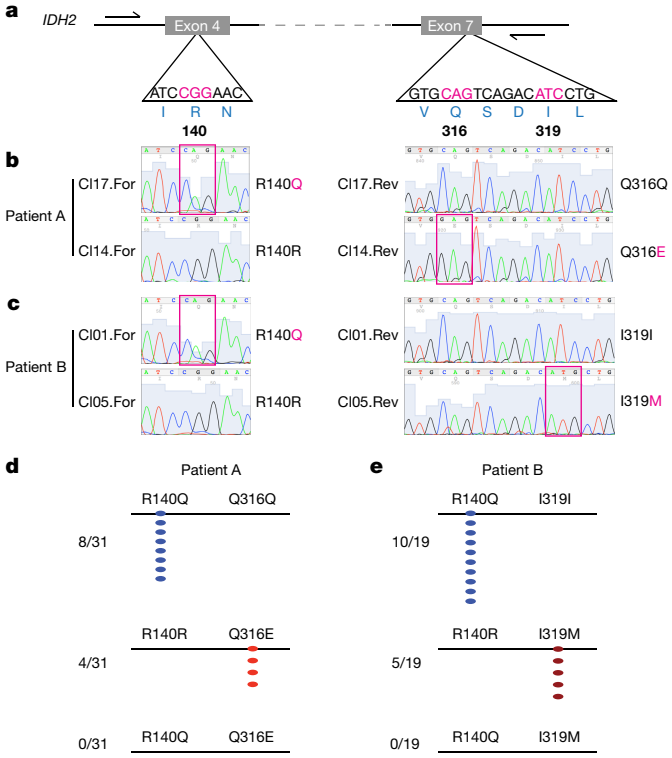


Fig. 2 | Second-site mutations in *IDH2* occur on the allele without the neomorphic R140Q mutation. a, Schematic of the *IDH2* locus (ENSG00000182054|CCDS10359), highlighting the nucleotides that encode arginine 140 (R140), glutamine 316 (Q316), and isoleucine 319 (I319). Positions of sequencing primers are indicated by half-arrows. **b**, **c**, Examples of Sanger sequencing in the forward ('For') and reverse ('Rev') direction from two clones ('Cl') for patient A (**b**) and patient B (**c**). Magenta boxes highlight the somatic mutations. **d**, **e**, Summary of Sanger sequencing results for patient A (**d**) and patient B (**e**), demonstrating that the R140Q mutations and the Q316E (**d**) or I319M (**e**) mutations do not occur on the same allele.

(Fig. 1d). Treatment with enasidenib led to a decline in leukaemic blasts in the bone marrow (Fig. 1a and Extended Data Fig. 1a), a reduction in plasma 2HG concentration to less than 1 μM (Fig. 1c), and normalization of the absolute neutrophil count (Fig. 1b). The variant allele frequency (VAF) for $IDH2^{R140Q}$ remained in the 20–50% range (Fig. 1d), consistent with the observation that enasidenib promotes differentiation of *IDH2*-mutant blasts[8,11,12]. After 9 months of therapy, the patient developed recurrent neutropenia and reappearance of leukaemic blasts consistent with disease progression (Fig. 1a, b and Extended Data Fig. 1a). Importantly, the plasma 2HG concentration had increased despite continuous treatment with enasidenib (Fig. 1c).

The second patient presented with AML that was refractory to cytarabine-based induction chemotherapy and subsequent therapy with an investigational DOT1L inhibitor (Fig. 1e). The patient had an increased blast count, neutropenia, and an increased plasma 2HG concentration of more than 5 μM (Fig. 1e–g and Extended Data Fig. 1b). Next-generation sequencing of bone marrow cells demonstrated the presence of an $IDH2^{R140Q}$ mutation (Fig. 1h). Treatment with enasidenib induced a considerable reduction in the peripheral blast count, an increase in the absolute neutrophil count, and a decrease in the plasma 2HG concentration to less than 1 μM (Fig. 1e–g and Extended Data Fig. 1b). The VAF for $IDH2^{R140Q}$ remained in the 15–40% range (Fig. 1h). After 6 months of continuous treatment with enasidenib, the patient developed progressive disease with a rise in leukaemic blasts (Fig. 1e and Extended Data Fig. 1b) and increased plasma 2HG levels (Fig. 1g).

We hypothesized that clinical resistance to enasidenib in the setting of a progressive increase in 2HG levels might be due to acquisition of a new somatic mutation that mediates drug resistance. In both cases, we identified new mutations in the *IDH2* gene at the time of acquired resistance, missense mutations that resulted in a substitution of glutamine 316 with glutamate (Q316E) in the first patient and substitution of isoleucine 319 with methionine (I319M) in the second patient (Fig. 1d, h). Droplet digital PCR (ddPCR) (Extended Data Table 1) demonstrated the secondary *IDH2* mutations were not detectable before treatment with enasidenib. Thus, the secondary *IDH2* mutations were either acquired during the course of treatment or originated in a rare subclone that was below the limit of detection of the ddPCR assay.

The identification of second-site mutations in *IDH2* in the setting of acquired resistance to enasidenib therapy raised the possibility of a resistance mechanism similar to those previously reported for tyrosine kinases such as BCR-ABL and EGFR[13–16]. The Q316E and I319M mutations are encoded in exon 7 of the *IDH2* gene, whereas the neomorphic R140Q mutation is encoded upstream in exon 4 (Fig. 2a). To determine the allelic conformation of the different *IDH2* mutations, we performed long-range PCR amplification of genomic DNA spanning exons 4–7 of *IDH2* followed by subcloning and sequence analysis of individual clones (Fig. 2a–c). In the first patient, all clones with the $IDH2^{R140Q}$ mutation were wild type at position Q316 (Fig. 2b, d), whereas all clones with the $IDH2^{Q316E}$ mutation were wild type for R140 (Fig. 2b, d). We observed analogous results for the second patient, such that the $IDH2^{I319M}$ and $IDH2^{R140Q}$ mutations were observed exclusively in different clones (Fig. 2c, e). These data demonstrate that acquired resistance to enasidenib was associated with emergence of second-site mutations *in trans* on the *IDH2* allele without the neomorphic R140Q mutation.

To investigate the potential significance of the Q316E and I319M mutations in IDH2, we mapped the mutations at Q316 and I319 to the recently published structure of an IDH2 dimer bound by enasidenib (Fig. 3a; RCSB Protein Data Bank (PDB) code 5I96)[9]. Q316 and I319 are located in the IDH2 dimer interface and are key residues that interact with enasidenib[9] (Extended Data Fig. 2). Structural modelling predicted that the Q316E mutation disrupts hydrogen bonding with enasidenib (Fig. 3b), whereas the I319M mutation creates steric hindrance that would impede binding of enasidenib (Fig. 3c). Even though the dimer interface is symmetrical and enasidenib is not, identical residues on either side of the interface can make different, but important, interactions with the drug (Fig. 3a and Extended Data Fig. 2), allowing second-site mutations at the interface to function *in trans* (and potentially also *in cis*, see below).

We evaluated the effect of the Q316E and I319M mutations on IDH2 enzymatic function alone and *in trans* with the R140Q mutation. Expression of the Q316E or I319M mutations in Ba/F3 hematopoietic cells did not result in increased 2HG production, in contrast to the known effect of the R140Q mutation on neomorphic IDH2 function (Fig. 3d). Enasidenib dose-dependently reduced 2HG levels in Ba/F3 cells that co-expressed wild-type and R140Q-mutant IDH2 *in trans*, consistent with inhibition of mutant IDH2 enzymatic activity as previously described[9,11,12] (Fig. 3e and Extended Data Fig. 3a). By contrast, Ba/F3 cells that expressed IDH2(R140Q) concurrently with either IDH2(Q316E) or IDH2(I319M) *in trans* continued to produce high levels of 2HG when exposed to enasidenib (Fig. 3e and Extended Data Fig. 3a), although this could be partially inhibited by higher doses of drug (Extended Data Fig. 3b).

We next tested the effect of the Q316E and I319M mutations on the self-renewal of primary mouse haematopoietic stem/progenitor cells (HSPC) in which IDH2(R140Q) is expressed from the endogenous locus[17]. Enasidenib inhibited the serial-replating capacity of HSPCs that expressed both wild-type and R140Q mutant IDH2 *in trans*, consistent with inhibition of mutant IDH2-induced gain of self-renewal (Fig. 3f). However, primary mouse HSPCs that expressed IDH2(R140Q) in conjunction with Q316E or I319M *in trans* maintained serial replating in the presence of enasidenib (Fig. 3f). Similar findings were observed in a primary murine-derived leukaemia model, in which IDH2(R140Q) and the FLT3 internal tandem duplication (ITD) are expressed from

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

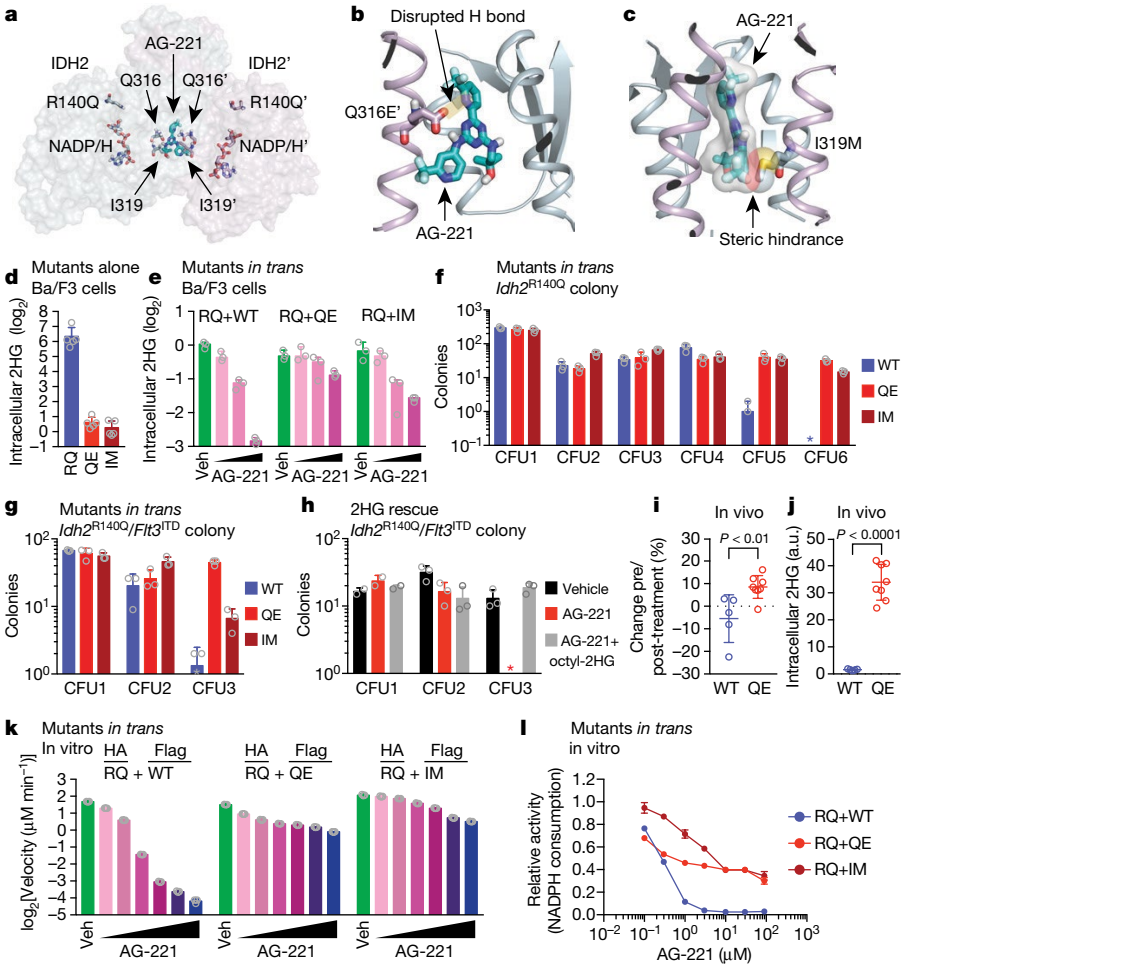


Fig. 3 | Second-site mutations in *IDH2* confer resistance to enasidenib *in trans*. **a**, Structure of the IDH2 dimer highlighting the binding pocket for enasidenib (AG-221; teal) at the dimer interface and the amino acids affected by second-site resistance mutations (Q316, I319; modelled from PDB code 5I96[9], see Methods). Second-site mutations are structurally distant from the catalytic active site containing the neomorphic R140Q mutation and NADP/H cofactor. **b**, Detailed view of the Q316E' mutation (apostrophe denotes second dimer subunit) showing loss of a hydrogen bond that normally forms between the amino side chain of Q316' and a nitrogen in the diaminotriazine ring of AG-221. **c**, Detailed view of the I319M mutation, demonstrating steric effects from the bulky side chain of methionine predicted to hinder binding by AG-221. **d**, Intracellular 2HG levels in Ba/F3 cells that express IDH2 mutations R140Q (RQ), Q316E (QE) or I319M (IM) via retroviral transduction. Values are relative to untransduced parental Ba/F3 cells. Data are mean ± s.e.m. for $n = 5$ cultures. **e**, Intracellular 2HG levels in Ba/F3 cells co-expressing IDH2 RQ plus wild-type (WT), QE or IM *in trans* and treated with vehicle ('Veh') or increasing doses of AG-221 (1, 10 or 100 nM). Data are mean ± s.e.m. for triplicate cultures. **f**, **g**, Serial-replating of primary HSPCs from IDH2(R140Q) (**f**) or IDH2(R140Q) and FLT3 ITD (**g**) mice that express wild-type, QE or IM IDH2 *in trans* and cultured in methylcellulose containing AG-221 at 50 nM. c.f.u., colony forming unit. Asterisk indicates value of 0. Data are mean ± s.e.m. for triplicate cultures. **h**, Serial-replating of primary HSPCs from $Idh2^{R140Q}/Flt3^{ITD}$ mice cultured in methylcellulose containing vehicle, AG-221 (50 nM) or AG-221 (50 nM) plus cell-permeable 2HG (octyl-2HG; 0.5 mM). Data are mean ± s.e.m. for duplicate (CFU1) or triplicate (CFU2/3) cultures. Asterisk indicates value of 0. **i**, **j**, Mice reconstituted with $Idh2^{R140Q}$ bone marrow HSPCs transduced with $IDH2^{WT}$ or $IDH2^{Q316E}$ were subjected to 2 (**i**) or 4 (**j**) weeks of treatment with enasidenib (40 mg kg^{-1} twice daily) and assessed for WT or QE allele frequencies before and after treatment (**i**) or intracellular 2HG levels in bone marrow mononuclear cells (**j**). See Methods. Data are mean ± s.e.m. for $n = 5$ WT and $n = 8$ QE mice. $P = 0.008$ (**i**) or $P = 4 \times 10^{-7}$ (**j**) by two-tailed *t*-test. **k**–**l**, In vitro enzyme assays measuring absolute velocity (**k**) and relative activity (**l**) of NADPH-dependent reduction of α-ketoglutarate (αKG) by haemagglutinin (HA)-precipitated IDH2 dimers purified from cells co-expressing IDH2 HA–RQ + Flag–WT, HA–RQ + Flag–QE, or HA–RQ + Flag–IM (see Methods). Reactions contained purified enzyme (10 μg ml^{-1}), NADPH (0.3 mM), αKG (5 mM) and AG-221 at 0.1, 0.3, 1, 3, 10 and 30 μM (**k**) or indicated concentrations (**l**). Data for **k**–**l** are mean ± 95% confidence intervals for triplicate reactions. Results are representative of ≥3 (**d**, **e**, **k**–**l**), 2 (**f**–**h**), or 1 (**i**, **j**) independent experiments. See also Extended Data Figs. 2–4.

their endogenous loci[17] (Fig. 3g and Extended Data Fig. 3c, d). Cell-permeable 2HG exposure was sufficient to confer enasidenib-resistant serial-replating to $Idh2^{R140Q}/Flt3^{ITD}$-mutant HSPC (Fig. 3h), supporting the hypothesis that restored 2HG production contributes to the resistance mediated by the IDH2 Q316E and I319M mutations.

We determined the effect of expressing IDH2(Q316E) *in trans* on response to enasidenib in vivo (Fig. 3i, j). Bone marrow HSPCs expressing IDH2(R140Q) were transduced with either $IDH2^{WT}$ or $IDH2^{Q316E}$ and used to reconstitute the hematopoietic system of recipient mice. Recipient mice were treated for 2 weeks with enasidenib, and $IDH2^{WT}$ and $IDH2^{Q316E}$ mutant allele frequencies were assessed before and after treatment. Enasidenib treatment reduced the proportion of $IDH2^{WT}$-transduced cells, whereas the frequency of $IDH2^{Q316E}$-transduced cells increased despite enasidenib treatment (Fig. 3i and Extended Data Fig. 3e; $P = 0.008$), consistent with a fitness advantage of $IDH2^{Q316E}$-transduced mutant cells during enasidenib therapy (see Fig. 1d). Similarly, the expression of the IDH2(Q316E) mutation *in trans* conferred production of 2HG by $IDH2^{R140Q}$-mutant cells in vivo that was resistant to enasidenib treatment (Fig. 3j; $P = 4 \times 10^{-7}$), consistent with the in vitro studies (Fig. 3e–g).

To rule out indirect cellular effects by which the Q316E and I319M mutants might induce resistance to IDH2 inhibitors, we purified IDH2

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

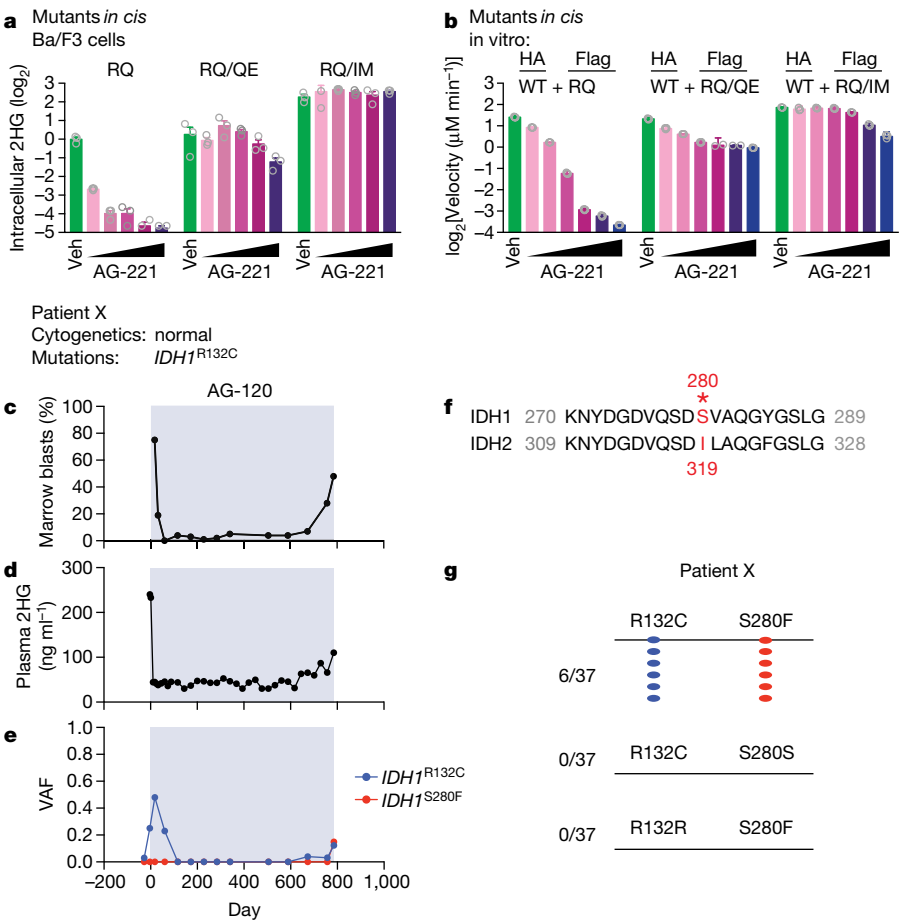


Fig. 4 | Second-site mutations *in cis* can confer resistance to IDH inhibitors. a, Intracellular 2HG levels in Ba/F3 cells transduced with $IDH2^{R140Q}$ (RQ), *in cis* double-mutant $IDH2^{R140Q/Q316E}$ (RQ/QE) or *in cis* double-mutant $IDH2^{R140Q/I319M}$ (RQ/IM) and treated with vehicle or increasing doses of AG-221 (10, 50, 100, 500 or 1,000 nM). Data are mean ± s.e.m. for triplicate cultures. **b**, In vitro enzyme assays measuring absolute velocity of NADPH-dependent reduction of αKG by HA-precipitated wild-type IDH2 dimerized with RQ alone or RQ plus *in cis* second-site mutations (RQ/QE or RQ/IM). See Extended Data Fig. 5. Reactions contained purified enzyme (7.5 μg ml^{-1}), NADPH (0.3 mM), αKG (5 mM), and either vehicle or increasing doses of AG-221 (0.1, 0.3, 1, 3, 10 or 30 μM). Data are mean ± 95% confidence intervals for triplicate reactions (duplicate reactions for WT + RQ/QE AG-221 3 μM and 30 μM). **c**–**e**, Clinical and laboratory features for patient X in relation to treatment with mutant IDH1 inhibitor ivosidenib (AG-120; blue box), including bone marrow blast percentage (**c**), plasma 2HG concentration (**d**), and VAF for mutations identified by targeted next-generation sequencing of bone marrow cells (**e**). **f**, Alignment of IDH1 and IDH2 protein sequences demonstrating that S280 of IDH1 corresponds to I319 of IDH2. **g**, Summary of Sanger sequencing results from patient X demonstrating that the $IDH1^{R132C}$ neomorphic mutation and the $IDH1^{S280F}$ mutation occur *in cis* on the same allele.

enzymatic dimers constituted of R140Q–wild type, R140Q–Q316E, or R140Q–I319M and performed in vitro enzymatic reactions (Fig. 3k, l and Extended Data Fig. 4a, b). We did not observe a difference in the ability of the Q316E and I319M mutants to form dimers compared to wild-type IDH2 (Extended Data Fig. 4c–e). In vitro enzymatic reactions with purified IDH2 dimers, α-ketoglutarate and NADPH demonstrated that each complex consumed NADPH and produced 2HG at similar rates, although R140Q–I319M dimers catalysed the reaction faster (Fig. 3k and Extended Data Fig. 4f). The activity of purified R140Q–wild-type enzyme complexes was inhibited by enasidenib in a dose-dependent manner, whereas purified R140Q–Q316E and R140Q–I319M enzyme dimers retained substantive activity in the presence of enasidenib (Fig. 3k, l and Extended Data Fig. 4f). These data demonstrate that although the Q316E and I319M mutations do not affect wild-type IDH2 enzymatic function, they can cooperate with the R140Q mutation *in trans* to maintain 2HG production and promote disease progression in the presence of enasidenib.

Although we observed the $IDH2^{Q316E}$ and $IDH2^{I319M}$ mutations *in trans* to the $IDH2^{R140Q}$ mutation, it is possible that these mutations might also have the capability to induce enasidenib resistance when acquired *in cis* to $IDH2^{R140Q}$. We therefore tested the possibility that these resistance mutations might confer resistance to enasidenib when present in the same allele as $IDH2^{R140Q}$. Consistent with the effect observed *in trans*, the expression of the $IDH2^{Q316E}$ or $IDH2^{I319M}$ mutations *in cis* with $IDH2^{R140Q}$ in Ba/F3 cells (Fig. 4a and Extended Data Fig. 5a) allowed for 2HG production that was not inhibited by enasidenib. In vitro enzymatic reactions with purified IDH2 dimers comprised of wild-type–R140Q/Q316E and wild-type–R140Q/I319M demonstrated that the enzymatic activity of these complexes was resistant to enasidenib (Fig. 4b and Extended Data Fig. 5b–g).

The experimental findings suggested that second-site mutations of *IDH2 in cis* also might mediate clinical resistance to enasidenib. Analysis of 14 patients with de novo resistance and 7 additional patients with acquired resistance to enasidenib did not identify any additional patients with second-site *IDH2* mutations in either allele (Extended Data Table 2). Notably, we identified one patient with $IDH1^{R132C}$-mutant AML who developed acquired clinical resistance to the mutant IDH1 inhibitor ivosidenib (AG-120) with a recurrent increase in blood levels of 2HG (Fig. 4c, d). In this patient, we identified a second-site mutation in *IDH1* that resulted in substitution of serine 280 with phenylalanine (S280F) (Fig. 4e). S280 in IDH1 is paralogous to I319 in IDH2 (Fig. 4f). To determine the allelic conformation of the *IDH1* mutations, we performed long-range PCR amplification of genomic DNA that encompassed both mutations followed by subcloning and

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

sequence analysis of individual clones. All clones with the *IDH1*^{R132C} mutation also harboured the *IDH1*^{S280F} mutation, demonstrating that the putative *IDH1*^{S280F} resistance mutation was acquired *in cis* with the neomorphic *IDH1*^{R132C} mutation (Fig. 4g).

In summary, we have identified a mechanism of acquired clinical resistance to a molecularly targeted therapy, in which a second-site mutation on the wild-type allele can cooperate with a heterozygous gain-of-function mutation on the other allele to induce therapeutic resistance. In theory, *in trans* resistance mutations may occur in other homomultimeric targets such as receptor tyrosine kinases. The possibility that *in trans* mutations of EGFR cysteine 797 (C797) could contribute to EGFR inhibitor resistance in cell lines has been reported, but it has not been observed to occur in vivo in the clinical context[18]. It remains unclear whether there may be a selective advantage to acquiring the *IDH2*^{Q316E} and *IDH2*^{I319M} mutations *in trans* versus *in cis*; however, we would predict that *in cis* mutations may be identified as an alternative mechanism of acquired resistance to mutant IDH2 inhibitors. Consistent with this hypothesis, we identified a patient with acquired resistance to IDH1 inhibition who acquired an analogous *IDH1* second-site mutation *in cis*. The identification of the IDH dimer-interface mutations credentials the importance of IDH as a therapeutic target and underscores the crucial role of 2HG in the pathogenesis of *IDH*-mutant malignancies.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0251-7

Received: 4 August 2017; Accepted: 2 May 2018;
Published online: 27 June 2018

1. Losman, J. A. & Kaelin, W. G. Jr. What a difference a hydroxyl makes: mutant IDH, (*R*)-2-hydroxyglutarate, and cancer. *Genes Dev.* **27**, 836–852 (2013).
2. Ley, T. J. et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
3. Ward, P. S. et al. The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting α-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **17**, 225–234 (2010).
4. Gross, S. et al. Cancer-associated metabolite 2-hydroxyglutarate accumulates in acute myelogenous leukemia with isocitrate dehydrogenase 1 and 2 mutations. *J. Exp. Med.* **207**, 339–344 (2010).
5. Losman, J. A. et al. (*R*)-2-hydroxyglutarate is sufficient to promote leukemogenesis and its effects are reversible. *Science* **339**, 1621–1625 (2013).
6. Figueroa, M. E. et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567 (2010).
7. Kats, L. M. et al. Proto-oncogenic role of mutant IDH2 in leukemia initiation and maintenance. *Cell Stem Cell* **14**, 329–341 (2014).
8. Chen, C. et al. Cancer-associated IDH2 mutants drive an acute myeloid leukemia that is susceptible to Brd4 inhibition. *Genes Dev.* **27**, 1974–1985 (2013).
9. Yen, K. et al. AG-221, a first-in-class therapy targeting acute myeloid leukemia harboring oncogenic *IDH2* mutations. *Cancer Discov.* **7**, 478–493 (2017).
10. Wang, F. et al. Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. *Science* **340**, 622–626 (2013).
11. Stein, E. M. et al. Enasidenib in mutant *IDH2* relapsed or refractory acute myeloid leukemia. *Blood* **130**, 722–731 (2017).
12. Amatangelo, M. D. et al. Enasidenib induces acute myeloid leukemia cell differentiation to promote clinical response. *Blood* **130**, 732–741 (2017).
13. Gorre, M. E. et al. Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science* **293**, 876–880 (2001).
14. Kobayashi, S. et al. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **352**, 786–792 (2005).
15. Pao, W. et al. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.* **2**, e73 (2005).
16. Choi, Y. L. et al. EML4-ALK mutations in lung cancer that confer resistance to ALK inhibitors. *N. Engl. J. Med.* **363**, 1734–1739 (2010).
17. Shih, A. H. et al. Combination targeted therapy to disrupt aberrant oncogenic signaling and reverse epigenetic dysfunction in *IDH2*- and *TET2*-mutant acute myeloid leukemia. *Cancer Discov.* **7**, 494–505 (2017).
18. Niederst, M. J. et al. The allelic context of the C797S mutation acquired upon treatment with third-generation EGFR inhibitors impacts sensitivity to subsequent treatment strategies. *Clin. Cancer Res.* **21**, 3924–3933 (2015).

Acknowledgements We thank members of the Levine and Thompson laboratories for discussions. We thank W.K. Chatila and N. Schultz for assistance with data deposition. A.M.I. is supported by the NIH/NCI (K08 CA201483), Leukemia & Lymphoma Society (3356-16), Burroughs Wellcome Fund (1015584), Susan & Peter Solomon Divisional Genomics Program, Steven A. Greenberg Fund, and Cycle for Survival. A.H.S. is supported by the NIH/NCI (K08 CA181507) and Leukemia & Lymphoma Society. The work was also supported, in part, by the Conquer Cancer Foundation of ASCO (A.M.I., A.H.S. and J.T.), the Leukemia & Lymphoma Society Specialized Center of Research Program (7011-16; A.M.I. and C.B.T.), a Translational and Integrative Medicine Research Fund (TIMRF) grant (A.H.S. and E.M.S.), the American Association for Cancer Research (J.T.), the American Society of Hematology/Robert Woods Johnson Foundation (J.T.), and grants from the NIH, including R01 CA168802-02 (C.B.T.), R35 CA197594-01A1 (R.L.L.), U54 OD020355 (R.L.L.), and the Memorial Sloan Kettering Cancer Center Support Grant (NIH P30 CA008748) including a supplement to R.L.L., C.B.T. and A.H.S. We acknowledge the use of the Integrated Genomics Operation Core, funded by the Memorial Sloan Kettering Cancer Center Support Grant (NIH P30 CA008748), Cycle for Survival, and the Marie-Josée and Henry R. Kravis Center for Molecular Oncology.

Reviewer information *Nature* thanks J. Cortes and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions A.M.I., A.H.S., C.B.T., R.L.L. and E.M.S. conceived the project, designed the experiments, analysed the data, and wrote the manuscript. A.M.I., A.H.S., B.W. and A.N. performed the experiments with technical assistance from F.M.C., N.T., V.D., H.L. and J.R.C. M.P., C.F., J.T. and M.S.T. assisted with management of clinical data and specimens. M.E.A. and M.R. assisted with pathological assessment of biospecimens. A.S.R., S.K.A., G.A.P. and J.D.C. performed the structural modelling. N.F. and E.P. performed the mutational analysis. B.W., S.C., Z.D.K. and S.A.B. assisted with identification and analysis of the IDH1-mutant leukaemia. All authors read and approved the manuscript.

Competing interests C.B.T. is a founder of Agios Pharmaceuticals and a member of its scientific advisory board. He also serves on the board of directors of Merck and Charles River Laboratories. R.L.L. is on the Supervisory Board of Qiagen. J.D.C. is a member of the scientific advisory board of Schrödinger. B.W., S.C., Z.D.K. and S.A.B. are employees of Agios Pharmaceuticals, Inc.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0251-7.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0251-7.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to C.B.T., R.L.L. or E.M.S.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Clinical specimens. The patients described were enrolled on the phase I/II studies NCT02074839 or NCT01915498[11]. Enrolment was open to all patients with relapsed or refractory AML with a mutation in *IDH1* (NCT02074839) or *IDH2* (NCT01915498) identified locally and confirmed centrally. Patients were required to be 18 years or older at the time of study entry. Both men and women were enrolled on the studies. Patients were required to have a performance status of 2 or better and adequate organ function as defined in the study protocols. Clinical data, blood and bone marrow samples from patients with AML were obtained after receiving written informed consent from patients. Approval was obtained from the Institutional Review Board at each institution participating in these clinical trials. Additional consent was obtained from participants at Memorial Sloan Kettering Cancer Center with analyses performed on the institutional biobanking protocol approved by the Institutional Review Board. Patient biospecimens were anonymized by creating unique identifiers with no associated protected health information (PHI) and keeping the key on a password-protected server. Data collection and research was performed in compliance with all relevant ethical regulations for human research participants. Absolute neutrophil counts and blast percentages were determined by standard clinical assays. Next-generation sequencing and determination of VAF were performed as previously described[12,19]. *IDH1*-mutant AML samples were assessed by the FoundationOne Heme panel as previously described[12,20].

Allele-specific sequencing and PCR assays. Genomic DNA was isolated from bone marrow mononuclear cells using the QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's instructions. The following primers were used to amplify the *IDH2* genomic region spanning exon 4 and exon 7: forward primer 5′-ATTCCTGTGCCCTCCTTTCT-3′; reverse primer 5′-CAG AGCCCACACATTTGCAC-3′. The ~2 kb PCR products were subcloned into a TA-cloning plasmid, followed by bacterial transformation, selection of individual bacterial colonies, isolation of plasmid DNA, and Sanger sequencing in the forward and reverse direction using the same primers. The *IDH1* genomic region was cloned and sequenced in a similar matter with the following primers: forward 5′-ACCAACGACCAAGTCACCAA-3′; reverse 5′-CCCTGGAATGACCCTGTTCC-3′. *IDH2* qPCR was performed with the following primers: R140R forward 5′-AGTCCCAATGGAACTATCCG-3′, R140Q forward 5′-AGTCCCAATGGAACTATCCA-3′, R140R/Q Reverse 5′-GGGGTGAAGACCATTTTGAA-3′, and with SYBR green reagent (Affymetrix). cDNA was prepared using the Verso cDNA synthesis kit (ThermoFisher). For ddPCR, an assay specific for each mutation was designed and ordered through Biorad. All reactions were performed on a QX200 ddPCR System (Biorad) and evaluated in technical duplicates. Reactions were partitioned into a median of ~16,000 droplets per well using the QX200 droplet generator and run on a 96-well thermal cycler. Plates were then analysed with the QuantaSoft v1.7 to assess the number of droplets positive for mutant or wildtype DNA.

Cell culture and DNA constructs. Adherent 293T cells (purchased from ATCC) were maintained at low passage number in high glucose DMEM with 10% FBS, glucose 25 mM, glutamine 4 mM, penicillin 100 U ml^{-1}, and streptomycin 100 $\mu g\,ml^{-1}$ and split every 2–3 days before reaching confluence. Suspension Ba/F3 cells (purchased from DSMZ) were maintained in RPMI with 10% FBS, mIL-3 3 ng ml^{-1}, penicillin 100 U ml^{-1}, and streptomycin 100 $\mu g\,ml^{-1}$. Cell lines were authenticated by Short Tandem Repeat (STR) profiling. Cell lines repeatedly tested negative for mycoplasma throughout the experimental period. Primary mouse HSPCs were isolated, transduced, and cultured in methylcellulose as previously described[6]. Colony forming unit (CFU) assays were performed in M3434 methylcellulose (Stem Cell Technologies) in the presence of DMSO, enasidenib (AG-221; Agios) at 50 nM, or (2*R*)-octyl-α-hydroxyglutarate (octyl-2HG; Cayman) at 0.5 mM. Colonies were counted every 7–10 days and re-plated. *IDH2* DNA constructs with C-terminal HA or Flag tags were cloned by standard site-directed mutagenesis (Agilent) and Gibson Assembly (New England Biolabs) into pCDNA3.1 (Addgene), MSCV-IRES-GFP (Addgene), or MSCV-IRES-mCherry (Addgene) vectors and verified by Sanger sequencing. For transient transfection experiments, 293T cells were transfected using polyethylenimine[21]. For retroviral infection experiments, supernatant from 293T cells transfected with helper virus and plasmids was collected after 72 h, filtered and applied to Ba/F3 parental cells or primary mouse HSPCs overnight. Then, 48 h after infection, GFP^+ and/or $mCherry^+$ cells were sorted by flow cytometry. Sorted cells were expanded and used for drug treatment experiments with enasidenib.

Mouse experiments. All animal procedures were conducted in accordance with the Guidelines for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committees (IACUC) at Memorial Sloan Kettering Cancer Center. C57BL/6 $Idh2^{R140Q}$ and $Idh2^{R140Q}/Flt3^{ITD}$ mice were previously described[17]. Bone marrow transplant and treatment of mice were performed as previously described[17]. In brief, 8–16-week old C57BL/6 $Idh2^{R140Q}$ female donor mice were treated with 5-flurouracil 0.15 mg g^{-1}, and bone marrow cells isolated 5 days after treatment. Cells were then infected with retrovirus encoding wild-type or Q316E mutant IDH2. Positive mCherry cells marking transduction were FACS sorted and injected into lethally irradiated 8–16-week old C57BL/6 wild-type female recipient mice (950 cGy) with 1.5×10^5 support marrow cells. Engrafted mice were then used as donors for secondary transplants into recipients that were used for treatment studies without randomization ($n = 5$ for $IDH2^{WT}$ and $n = 8$ for $IDH2^{Q316E}$. Investigators were not blinded to treatment. Enasidenib was given at a dose of 40 mg kg^{-1} twice daily by oral gavage. After 2 weeks of drug treatment, allele burden was assessed through peripheral blood mCherry positivity in red blood cell-lysed samples. After 4 weeks of drug treatment, the mice were euthanized, and bone marrow was harvested from femur and tibia bones by centrifugation for 1 min at 6,000*g* in 200 μl RPMI 10% FBS. Red blood cells were lysed in ACK (ammonium-chloride-potassium) lysis buffer. Cells were counted, pelleted, and frozen before 2HG analysis.

Gel electrophoresis and western blotting. For denatured gel electrophoresis, cells were collected in 1× RIPA buffer (Cell Signaling), sonicated, centrifuged at 21,000*g* at 4 °C, and supernatants were collected. Cleared cell lysates were quantified by BCA assay (ThermoFisher) and normalized for total protein concentration. Samples were separated by SDS-PAGE, then either stained directly with Coomassie Blue Reagent or transferred to nitrocellulose membranes (Life Technologies), blocked in 5% milk prepared in TBS with 0.1% Tween 20 (TBST), incubated with primary antibodies overnight at 4 °C then horseradish peroxidase (HRP)-conjugated secondary antibodies (GE Healthcare; anti-mouse, NA931V, sheep, 1:5,000; anti-rabbit, NA934V, donkey, 1:5,000) for 1 h the following day. After incubation with ECL (ThermoFisher or GE Healthcare), digital imaging was performed using the Amersham Imager 600 (GE Healthcare) or film imaging was performed using the SRX-101A (Konica Minolta). Primary antibodies used included: anti-Flag (Sigma, F1804; clone M2; mouse; 1:1,000), anti-GAPDH (Cell Signaling Technology, 5174; clone D16H11; rabbit; 1:1,000), anti-HA (Cell Signaling Technology, 2367S; clone 6E2; mouse; 1:1,000), anti-IDH2 (Abcam, ab55271; no clone name; mouse; 1:1,000), and anti-vinculin (Cell Signaling Technology, 4650; no clone name; rabbit; 1:1,000). For native gel electrophoresis, cells were harvested in M-PER buffer (ThermoFisher) without sonication, IDH2 enzyme complexes were purified with Pierce Anti-HA Agarose (ThermoFisher), separated by NativePAGE (ThermoFisher), and either stained directly with Coomassie reagent or transferred for western blotting as described above.

Enzyme assays. HA-tagged IDH2 enzymes were purified from transfected 293T cells using Pierce Anti-HA Agarose (ThermoFisher) according to the manufacturer's instructions. HA-tagged enzymes were quantified by denatured gel electrophoresis with Coomassie staining in reference to a defined quantity of recombinant human IDH2 (Abcam, ab198092). Purified enzymes were used at 7.5–10 $\mu g\,ml^{-1}$ as indicated. NAPDH was used at 0.30 mM and α-ketoglutarate was used at 5 mM unless otherwise indicated. The enzyme reaction buffer consisted of HEPES 50 mM, NaCl 150 mM, $MgCl_2$ 20 mM, and BSA 0.01%. For NADPH consumption assays, reactions were conducted in UV-transparent 96-well plates (Corning) with reaction volumes of 200 μl. A SpectraMax Plus 384 Microplate Reader (Molecular Devices) was used to monitor the absorbance at 340 nm every 30 seconds throughout the course of the reaction. For each condition, the mean rate of NADPH consumption for triplicate control reactions without enzyme was subtracted from the rate of NADPH consumption for triplicate experimental reactions with enzyme. Reaction velocities were calculated using an extinction coefficient for NADPH at ε340 of 6,220 $M^{-1}\,cm^{-1}$ and pathlength of 0.56 cm for a 200 μl reaction volume in a standard 96-well plate.

Metabolite extraction and analysis. Metabolites were extracted with ice-cold 80:20 methanol:water containing 2 μM deuterated 2-hydroxyglutarate (D-2-hydroxyglutaric-2,3,3,4,4-d_5 acid; deuterated-2HG) as an internal standard. After overnight incubation at −80 °C, cell extract was collected, sonicated and centrifuged at 21,000*g* for 20 min at 4 °C to precipitate protein. Extracts were then dried in an evaporator (Genevac EZ-2 Elite). For gas chromatography–mass spectrometry (GC–MS), metabolites were resuspended by addition of 50 μl of methoxyamine hydrochloride (40 mg ml^{-1} in pyridine) and incubated at 30 °C for 90 min with agitation. Metabolites were further derivatized by addition of 80 μl of *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (MSTFA) plus 1% 2,2,2-trifluoro-*N*-methyl-*N*-(trimethylsilyl)-acetamide, chlorotrimethylsilane (TCMS; Thermo Scientific) and 70 μl of ethyl acetate (Sigma) and incubated at 37 °C for 30 min. Samples were diluted 1:2 with 200 μl of ethyl acetate, then analysed using an Agilent 7890A GC coupled to Agilent 5975C mass selective detector. The GC was operated in splitless mode with constant helium carrier gas flow of 1 ml min^{-1} and with a HP-5MS column (Agilent Technologies). The injection volume was 1 μl and the GC oven temperature was ramped from 60 °C to 290 °C over 25 min. Peaks representing compounds of interest were extracted and integrated using MassHunter vB.08.00 (Agilent Technologies) and then normalized to both the internal standard (deuterated-2HG) peak area and cell number or protein content as applicable. Ions used for quantification of metabolite levels were

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

2HG *m/z* 247 (confirmatory ion *m/z* 349) and deuterated-2HG *m/z* 252 (confirmatory ion *m/z* 354). Peaks were manually inspected and verified relative to known spectra for each metabolite. Absolute metabolite quantitation was performed using an external calibration curve with deuterated-2HG internal standard and the resulting concentrations corrected for the total cell volume extracted. For IDH1-mutant AML patients, plasma 2HG concentrations were determined using a qualified liquid chromatography–tandem mass spectrometry (LC–MS/MS) method with a lower limit of quantitation of 30 ng ml^{-1} as previously described[22].

Structural modelling. The structure of enasidenib (AG-221) bound to IDH2 was obtained from PDB code 5I96 retrieved from the RCSB[9]. This structure has a R140Q mutation in both homodimer subunits. Residues I170, L298, S300, V315, Q316, S317, D318, I319, L320 and I170', V294', S300', V315', Q316', S317', D318', I319', L320', and AG-221 had two conformations resolved with different occupancies; here, residues with an apostrophe denote second dimer subunit. Two models were constructed: one for the high-occupancy conformations of all residues, and another for the low-occupancy conformations of all residues. Point mutants of interest (Q316E, I319M, Q316E' and I319M') were introduced in the structures of each of the two models using the 'mutate residues' functionality in Maestro 11.2 (Schrödinger)23, and side-chain orientations were optimized using the 'predict side chains' functionality with default setting in Prime v4.8 (Schrödinger)[23–25]. This resulted in a total of eight single-mutant structures. Inspection of the mutated structures revealed potential differences in interactions between mutant and wild-type that could affect the binding of AG-221. These features were manually highlighted in figures produced using PyMOL v1.8.2.0 (Schrödinger; https://pymol.org/2/).

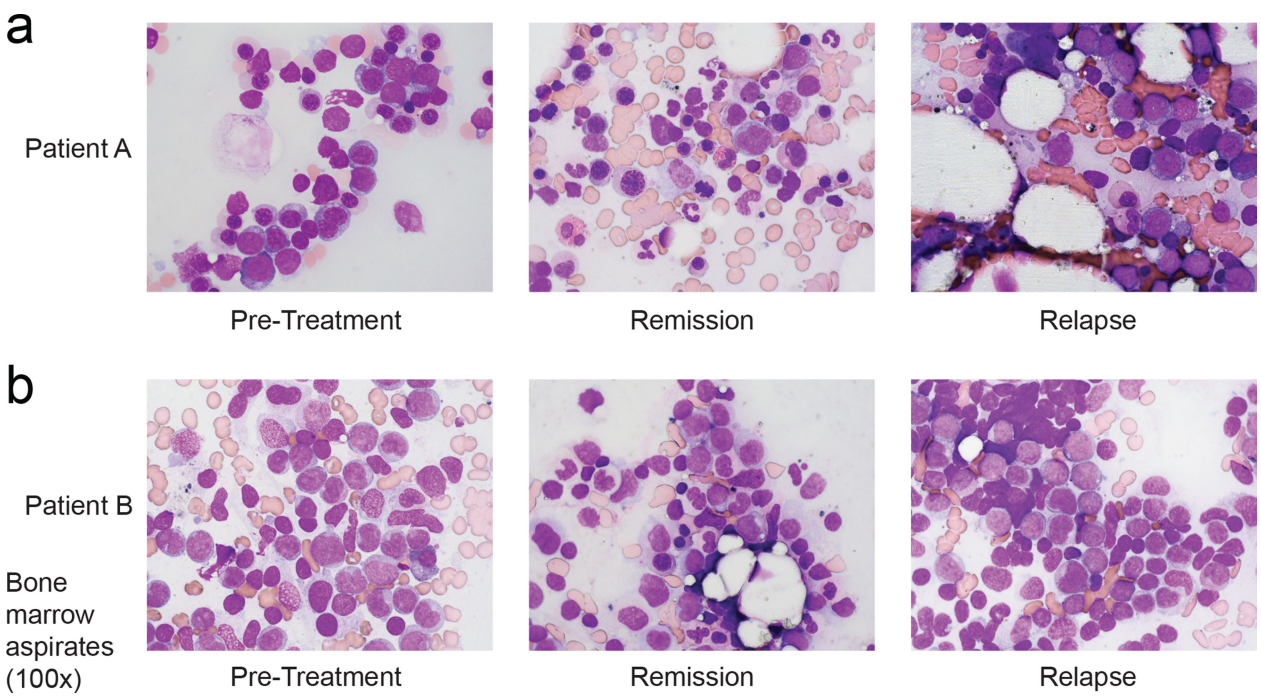
Statistics and reproducibility. Significance was determined by two-tailed Student's *t*-test comparing the indicated condition to the corresponding wild-type or control. Open circles are individual data points. For quantitative measurements, *n* is provided in the figure legends. All results have been independently replicated at least twice, with the exception of the in vivo experiments shown in Fig. 3i, j. No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. Source Data for all figures are provided with the paper. Uncropped versions of blots are provided in Supplementary Figs. 1–3. Structural modelling studies referenced PDB accession code 5I96. Images, settings and scripts for the structural modelling are available at https://doi.org/10.6084/m9.figshare.5966878. Sequencing data are deposited in the European Nucleotide Archive (ENA) under accession number PRJEB26337.

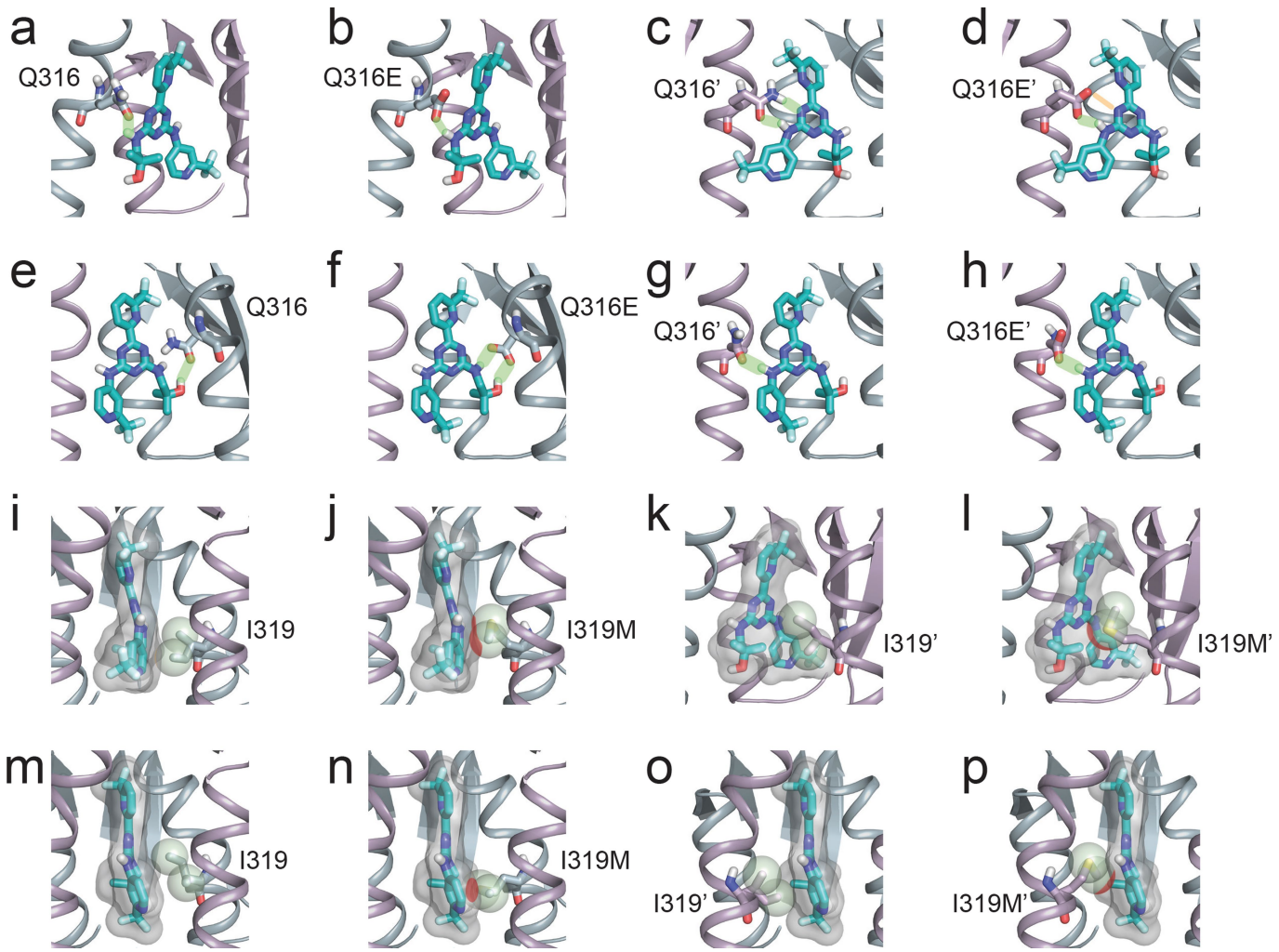
19. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
20. He, J. et al. Integrated genomic DNA/RNA profiling of hematologic malignancies in the clinical setting. *Blood* **127**, 3004–3014 (2016).
21. Yang, S., Shi, H., Chu, X., Zhou, X. & Sun, P. A rapid and efficient polyethylenimine-based transfection method to prepare lentiviral or retroviral vectors: useful for making iPS cells and transduction of primary cells. *Biotechnol. Lett.* **38**, 1631–1641 (2016).
22. Dang, L. et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
23. Schrödinger. *Schrödinger Release 2017-2: Maestro* (Schrödinger, LLC, New York, 2017).
24. Jacobson, M. P., Friesner, R. A., Xiang, Z. & Honig, B. On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.* **320**, 597–608 (2002).
25. Jacobson, M. P. et al. A hierarchical approach to all-atom protein loop prediction. *Proteins* **55**, 351–367 (2004).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



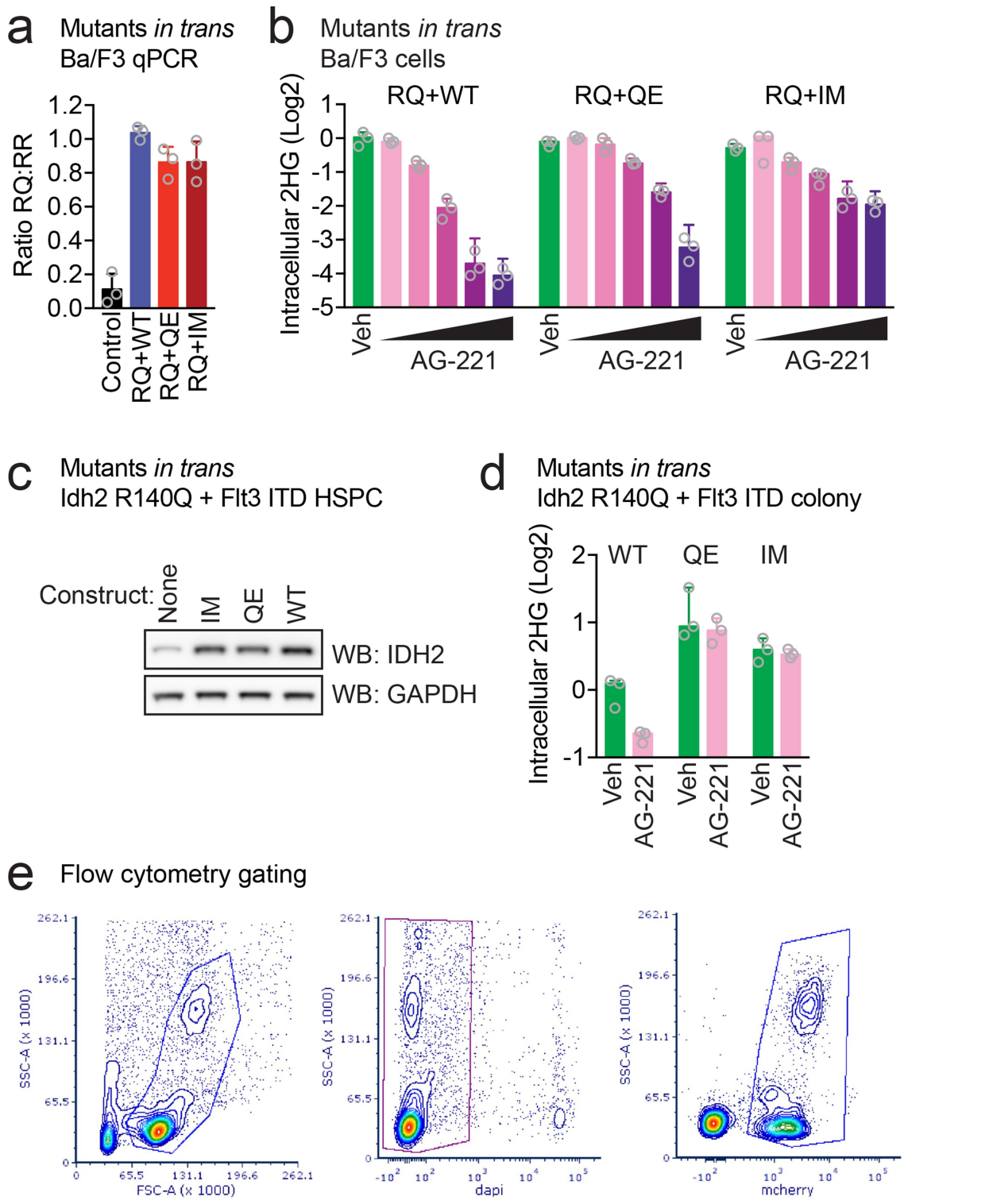
Extended Data Fig. 1 | Acquired clinical resistance to the mutant IDH2 inhibitor enasidenib (AG-221). a, **b**, Haematoxylin and eosin staining of bone marrow cells aspirated from patient A (**a**) and patient B (**b**) at indicated points in relation to treatment with AG-221. Remission images demonstrate decreased leukaemic blasts and increased myeloid differentiation that are reversed at the time of relapse. Images show 100× magnification. Images are representative fields of a single bone marrow aspiration performed at each time point.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



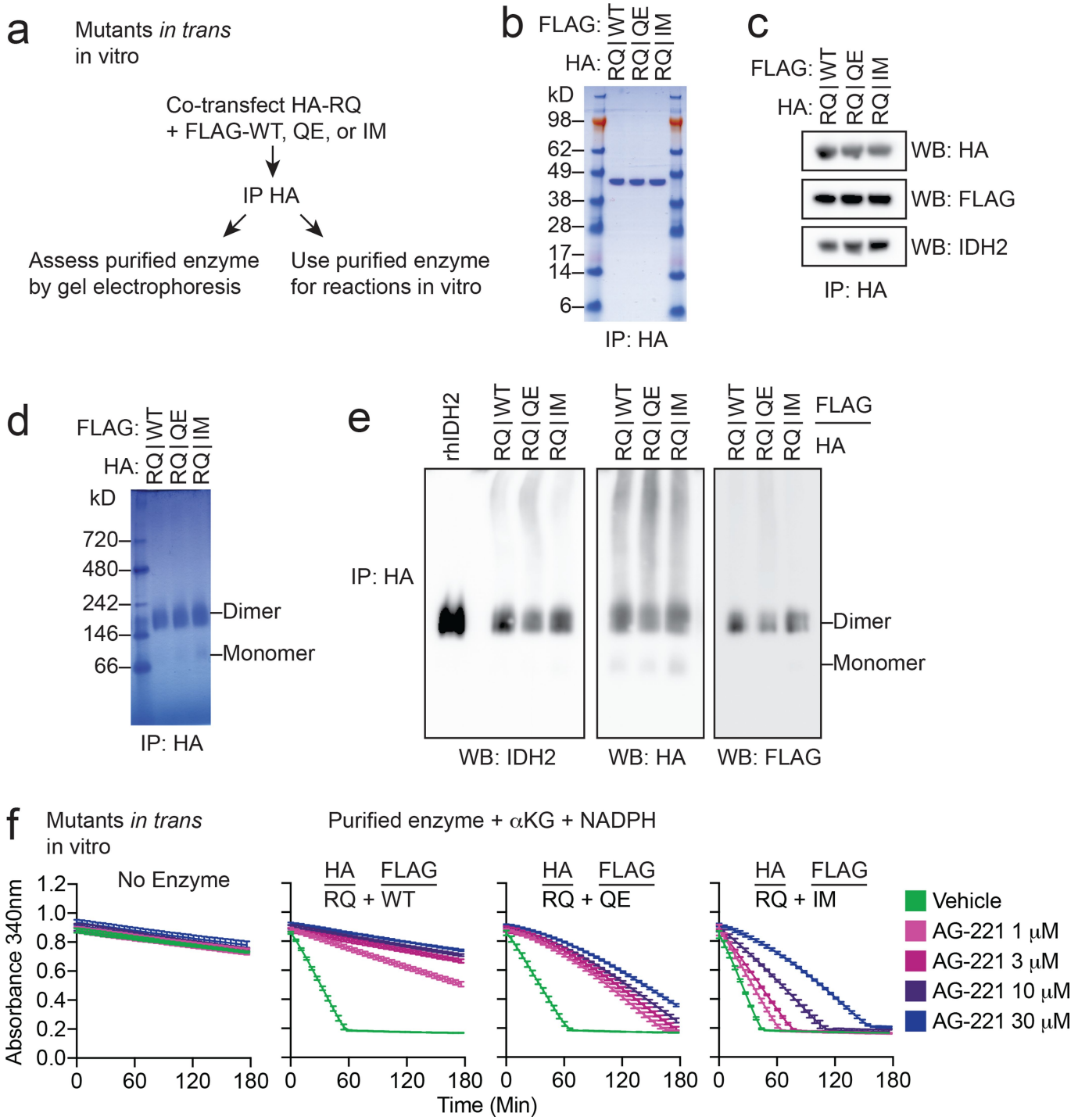
Extended Data Fig. 2 | Structures illustrating potential interactions between IDH2 second-site mutations and enasidenib. a–h, Detailed view of the interactions between wild-type IDH2 Q316 (**a**, **e**) and Q316' (**c**, **g**) or mutant IDH2(Q316E) (**b**, **f**) and IDH2(Q316E') (**d**, **h**) with AG-221 in the predicted dominant conformation (**a**–**d**) or a minor conformation (**e**–**h**). Hydrogen bonds are depicted in light green. Note the disrupted hydrogen bond (depicted as orange bar) in **d** resulting from the Q316E mutation in the IDH2' subunit. **i–p**, Detailed view of the interactions between wild-type IDH2 I319 (**i**, **m**) and I319' (**k**, **o**) or mutant IDH2(I319M) (**j**, **n**) and IDH2(I319M') (**l**, **p**) with AG-221 in the predicted dominant conformation (**i**–**l**) or a predicted minor conformation (**m**–**p**). The solvent-excluded surface of AG-221 is shown transparently in grey. The van der Waals radius of the Cδ1 and Cγ2 atoms of I319/I319' or the Sδ and Cε atoms of I319M/I319M' are depicted as spheres. Unfavourable steric interactions between AG-221 and these atoms are depicted in red. Throughout the figure, the IDH2 subunit is depicted in blue-grey, the IDH2' subunit in purple, and AG-221 in teal. Non-polar hydrogen atoms are not shown. White, red, blue and yellow portions of stick structures indicate hydrogen, oxygen, nitrogen and sulfur atoms, respectively. All models were based on the AG-221–IDH2 structure (PDB code 5I96)[9] (see Methods).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



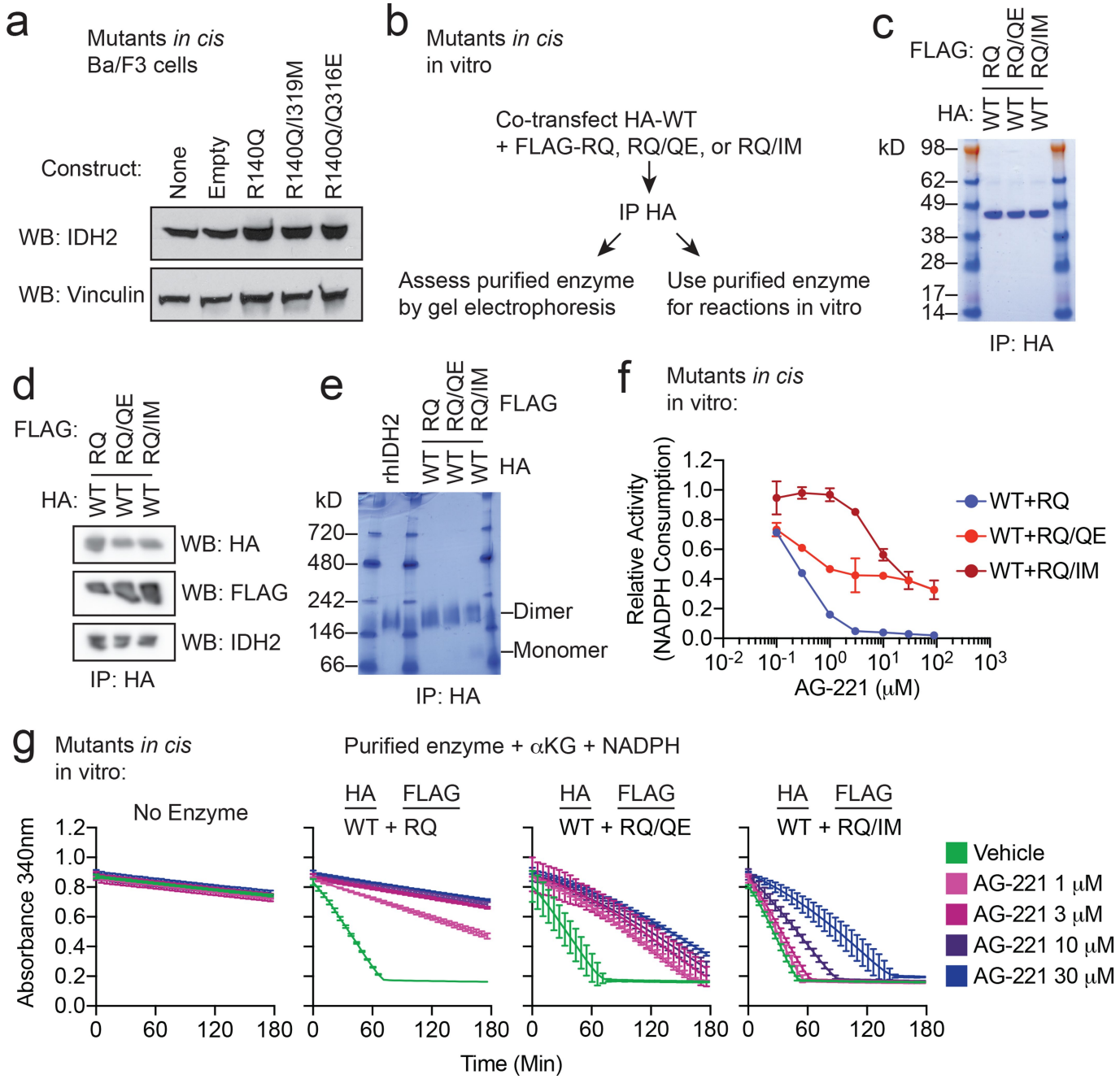
Extended Data Fig. 3 | Expression and activity of *in trans* IDH2 second-site mutations in haematopoietic cells. a, Allele-specific quantitative PCR (qPCR) showing similar expression of constructs in Ba/F3 cells co-transduced with $IDH2^{R140Q}$ (RQ) plus $IDH2^{WT}$ (WT), $IDH2^{Q316E}$ (QE), or $IDH2^{I319M}$ (IM) *in trans*. Control from $IDH2^{WT}$ human cell line (293T). Data are mean ± s.e.m. for triplicate reactions. **b**, Intracellular 2HG levels in Ba/F3 cells co-expressing RQ plus WT, QE or IM *in trans* and treated with vehicle or increasing doses of AG-221 (1 nM, 10 nM, 100 nM, 1 μM or 10 μM). Data are mean ± s.e.m. for triplicate cultures. **c**, Western blot showing IDH2 protein levels in primary HSPCs from $Idh2^{R140Q}/Flt3^{ITD}$ mice transduced with WT, QE or IM and untransduced control cells for comparison. GAPDH serves as a loading control. The same membrane was stripped and reprobed for western blots. **d**, Intracellular 2HG levels in primary HSPCs from $Idh2^{R140Q}/Flt3^{ITD}$ mice transduced with WT, QE or IM and collected from the first passage of methylcellulose cultures containing AG-221 at 50 nM. Data for are mean ± s.e.m. for triplicate cultures. **e**, Flow cytometry gating strategy for Fig. 3i. SSC-A, side scatter area; FSC-A, forward scatter area. DAPI is a viability dye. mCherry identifies retrovirally transduced cells. Results are representative of ≥2 (**a**–**d**) or 1 (**e**) independent experiments. For gel source data, see Supplementary Fig. 1.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 4 | Purification and activity of IDH2(R140Q) dimers with wild-type IDH2 or mutants Q316E or I319M *in trans*. **a**, Schematic of experimental approach: 293T cells were co-transfected with HA-tagged IDH2(R140Q) plus Flag-tagged wild-type, Q316E or I319M. After 2 days, cells were lysed and enzyme complexes were purified by HA-immunoprecipitation. Reactions were performed with purified enzyme, NADPH, αKG and varying doses of AG-221 as detailed in Fig. 3. **b**–**e**, Purity and dimerization of HA-precipitated enzymes were assessed by denatured SDS–PAGE with Coomassie staining (**b**), denatured SDS–PAGE with western blotting (**c**), native PAGE with Coomassie staining (**d**), or native PAGE with western blotting for the indicated proteins (**e**). Separate membranes were used for western blots. **f**, In vitro enzyme assays measuring rate of NADPH consumption of IDH2 dimers purified as in **b**–**e**. Reactions contained purified enzyme (10 μg ml^{-1}), NADPH (0.3 mM), αKG (5 mM) and AG-221 at indicated concentrations. Data are mean ± 95% confidence intervals for triplicate reactions. Results are representative of ≥3 (**b**–**d**, **f**) or 2 (**e**) independent experiments. For gel source data, see Supplementary Fig. 2.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 5 | Second-site *IDH2* mutations *in cis* can confer resistance to enasidenib. a, Western blot showing IDH2 expression in Ba/F3 cells transduced with the indicated constructs. Vinculin serves as a loading control. The same membrane was probed for both IDH2 and vinculin. These are the same cells as in Fig. 4a. **b**–**g**, Purification and enzymatic activity of IDH2 WT–R140Q dimers with or without *in cis* second-site mutations. **b**, Schematic of experimental approach: 293T cells were co-transfected with HA-tagged wild-type IDH2 plus Flag-tagged IDH2(R140Q), *in cis* double-mutant IDH2 R140Q/Q316E (RQ/QE) or *in cis* double-mutant IDH2 R140Q/I319M (RQ/IM). After 2 days, cells were lysed and IDH2 enzyme complexes were purified by HA-immunoprecipitation. **c**–**e**, Purity and dimerization of HA-precipitated enzymes were assessed by denatured SDS–PAGE with Coomassie staining (**c**), denatured SDS–PAGE with western blotting with the indicated antibodies (**d**), or native PAGE with Coomassie staining (**e**). Separate membranes were used for Western blots. **f**, **g**, In vitro enzyme assays measuring relative activity (**f**) and rate of NADPH consumption (**g**) by HA-precipitated IDH2 dimers. Reactions contained purified enzyme (7.5 μg ml^{-1}), NADPH (0.3 mM), αKG (5 mM), and vehicle or increasing doses of AG-221 (0.1, 0.3, 1, 3, 10 or 30 μM). Data are mean ± 95% confidence interval for triplicate reactions (duplicate reactions for WT–RQ/QE AG-221 3 μM and 30 μM). Results are representative of ≥3 independent experiments. For gel source data, see Supplementary Fig. 3.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | ddPCR for *IDH2* mutations in pre- and post-treatment samples

Patient A

Assay	Sample	Day	Droplet Count Mutant	Droplet Count WT	Ratio ([Mut]/[WT])
IDH2_Q316E	BM pre	-91	0	5081	0.00
IDH2_Q316E	BM pre	-14	0	2666	0.00
IDH2_Q316E	BM post	364	1089	1309	0.83
IDH2_R140Q	BM pre	-91	2413	3242	0.74
IDH2_R140Q	BM pre	-14	1254	1396	0.90
IDH2_R140Q	BM post	364	1232	1365	0.90

Patient B

Assay	Sample	Day	Droplet Count Mutant	Droplet Count WT	Ratio ([Mut]/[WT])
IDH2_I319M	BM pre	-32	0	8767	0.00
IDH2_I319M	PB pre	-11	0	11695	0.00
IDH2_I319M	BM post	465	800	3684	0.22
IDH2_R140Q	BM pre	-32	5086	5718	0.89
IDH2_R140Q	PB pre	-11	6270	6389	0.98
IDH2_R140Q	BM post	465	2337	2570	0.91

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 2 | Frequency of second-site *IDH2* mutations in AML patients treated with AG-221

Category	N (%)	Second-site IDH2 mutation (in subset)
Total treated on AG-221 protocol at MSKCC and assessed by next-gen sequencing	59 (100%)	0/59 (0%)*
De novo drug resistance	14/59 (24%)	0/14 (0%)
Acquired clinical resistance	9/59 (15%)	2/9 (22%)
• Normal blood 2HG	5/59 (8.5%)	0/5 (0%)
• Elevated blood 2HG	4/59 (6.8%)	2/4 (50%)

*Pre-treatment/earliest sample.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

https://doi.org/10.1038/s41586-018-0258-0

The helicase Ded1p controls use of near-cognate translation initiation codons in 5′ UTRs

Ulf-Peter Guenther[1,8], David E. Weinberg[2,3,4,8], Meghan M. Zubradt[2,5,8], Frank A. Tedeschi[1], Brittany N. Stawicki[1], Leah L. Zagore[1], Gloria A. Brar[6], Donny D. Licatalosi[1], David P. Bartel[3,4], Jonathan S. Weissman[2,5] & Eckhard Jankowsky[1,7]*

The conserved and essential DEAD-box RNA helicase Ded1p from yeast and its mammalian orthologue DDX3 are critical for the initiation of translation[1]. Mutations in DDX3 are linked to tumorigenesis[2–4] and intellectual disability[5], and the enzyme is targeted by a range of viruses[6]. How Ded1p and its orthologues engage RNAs during the initiation of translation is unknown. Here we show, by integrating transcriptome-wide analyses of translation, RNA structure and Ded1p–RNA binding, that the effects of Ded1p on the initiation of translation are connected to near-cognate initiation codons in 5′ untranslated regions. Ded1p associates with the translation pre-initiation complex at the mRNA entry channel and repressing the activity of Ded1p leads to the accumulation of RNA structure in 5′ untranslated regions, the initiation of translation from near-cognate start codons immediately upstream of these structures and decreased protein synthesis from the corresponding main open reading frames. The data reveal a program for the regulation of translation that links Ded1p, the activation of near-cognate start codons and mRNA structure. This program has a role in meiosis, in which a marked decrease in the levels of Ded1p is accompanied by the activation of the alternative translation initiation sites that are seen when the activity of Ded1p is repressed. Our observations indicate that Ded1p affects translation initiation by controlling the use of near-cognate initiation codons that are proximal to mRNA structure in 5′ untranslated regions.

To systematically analyse how Ded1p influences translation initiation in cells, we first examined how a mutation in the enzyme altered the spectrum of ribosome footprints in cells[7]. We used the *ded1-95* mutation (*Ded1p*^{T408I}), which reduces the affinity of Ded1p for RNA, diminishes RNA unwinding and confers a temperature-sensitive growth defect to the budding yeast *Saccharomyces cerevisiae*[8]. The mutation does not affect pre-mRNA splicing or ribosome biogenesis[9]. We performed ribosome profiling on wild-type and *ded1-95* strains before and after a temperature shift from 30 °C to 37 °C for 5 min (Extended Data Fig. 1a–h). The short time was chosen to minimize any secondary effects on ribosome footprints arising from broader translation defects.

At 30 °C, wild-type and *ded1-95* strains showed virtually indistinguishable RNA expression and translation profiles (Extended Data Fig. 1i, j). After the temperature shift, translation broadly decreased in *ded1-95*, compared to the wild type (Extended Data Fig. 1k–n). These observations indicate that Ded1p promotes translation initiation for most mRNAs, consistent with previous findings[10,11]. However, translation of a subset of mRNAs coding for proteins involved in gluconeogenesis, cell wall synthesis and transcripts encoding histones were less affected by Ded1p than other mRNAs (Extended Data Fig. 2a, b).

The fraction of ribosomes on 5′ untranslated regions (UTRs) markedly increased upon temperature shift in *ded1-95*, compared to the wild type (Fig. 1a, Extended Data Fig. 3a). The majority of mRNAs showed higher ribosome occupancy of the 5′ UTR in the *ded1-95* strain, which correlated with lower translation efficiency of the main open reading frame (ORF) (Fig. 1b, Extended Data Fig. 3b). To examine the link

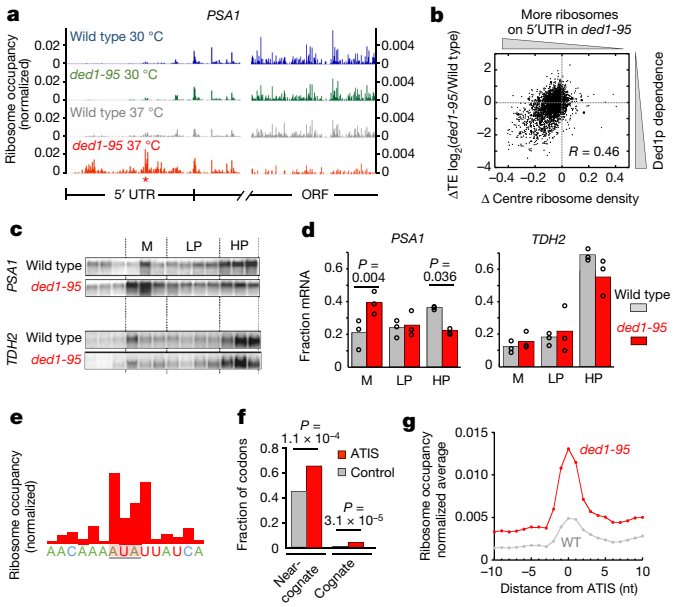


Fig. 1 | Defects in Ded1p activate alternative translation initiation sites. a, Ribosome profiling tracks of the 5′ UTR of *PSA1* mRNA for wild-type *DED1* and *ded1-95*, before and 5 min after a temperature shift. Bars mark the ribosome P-site, the star an alternative translation initiation site. Similar results were obtained in two independent experiments for each dataset. **b**, Correlation between change in translational efficiency (ΔTE) and change in the centre of ribosome density of wild-type *DED1*, compared to *ded1-95* ($n = 2{,}837$, 5 min, 37 °C). *R*, Pearson's correlation coefficient. **c**, Representative RNA blots of *PSA1* ($\log_2\Delta TE_{PSA1} = -2.1$) and *TDH2* ($\log_2\Delta TE_{TDH2} = 0.7$) after polysome fractionation for wild-type *DED1* and *ded1-95*, 5 min after temperature shift. M, 80S monosomes; LP, light polysomes; HP, heavy polysomes. Similar results were obtained in three independent experiments. **d**, Quantification of *PSA1* and *TDH2* RNA blots. Bars indicate the fraction of the mRNA in monosomes (M), light polysomes (LP), and heavy polysomes (HP). *P* values from a two-tailed *t*-test. **e**, Representative ribosome profiling track for a segment in the 5′ UTR of *PSA1* (indicated by the star in **a** in *ded1-95* (5 min, 37 °C)). The near-cognate initiation codon is highlighted. Similar results were obtained in two independent experiments **f**, Fraction of near-cognate and cognate initiation codons at sites with marked ribosome accumulation (red bars, ATISs, $n = 396$), and at randomly chosen control positions (grey) in 5′ UTRs in *ded1-95*. *P* values from a two-tailed *t*-test. **k**, Mean ribosome occupancy 10 nt 3′ and 5′ of high-confidence ATIS on 5′ UTRs (moving average of ±1 nt) for *ded1-95* (red) and wild-type *DED1* (5 min, 37 °C).

[1]Center for RNA Science and Therapeutics, School of Medicine, Case Western Reserve University, Cleveland, OH, USA. [2]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA, USA. [3]Howard Hughes Medical Institute Whitehead Institute for Biomedical Research, Cambridge, MA, USA. [4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Howard Hughes Medical Institute California Institute for Quantitative Biomedical Research, San Francisco, CA, USA. [6]Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA. [7]Department of Physics, Case Western Reserve University, Cleveland, OH, USA. [8]These authors contributed equally: Ulf-Peter Guenther, David E. Weinberg, Meghan M. Zubradt. *e-mail: exj13@case.edu

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

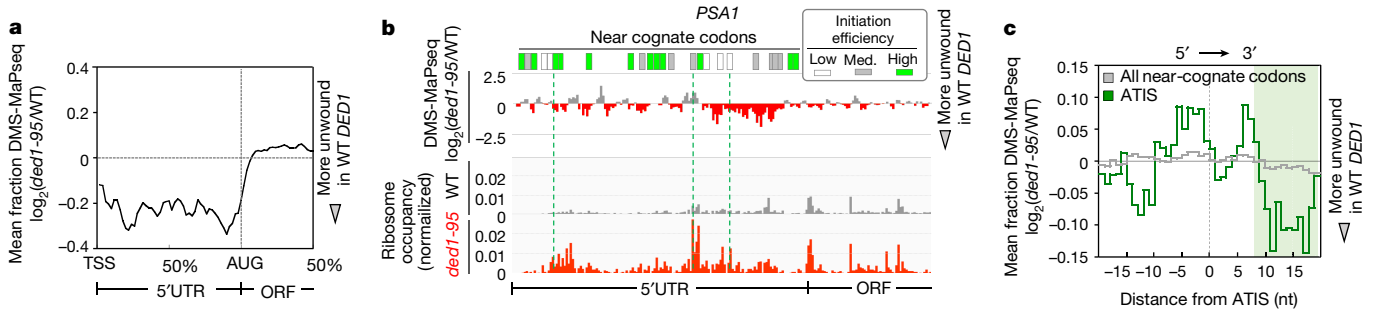


Fig. 2 | mRNA structure unwinding by Ded1p and ATIS activation. **a**, Metagene profile of mRNA unwinding by Ded1p on 5′ UTRs and the 5′ moiety of ORFs (moving average of ±1 nt). Similar results were obtained in two independent experiments. TSS, transcription start site; AUG, translation start site. **b**, Representative differential DMS mutational profiling with sequencing (DMS-MaPseq) track for the 5′ UTR of the *PSA1* mRNA (upper track, mRNA regions unwound in wild-type *DED1* (WT) marked by red bars, the more negative the value, the stronger the unwinding). Similar results were obtained in two independent experiments. For comparison, ribosome profiling traces of the 5′ UTR of *PSA1* mRNA for wild-type *DED1* and *ded1-95* are shown. Near-cognate codons are colour-coded according to their initiation efficiency, as indicated. Green lines mark activated ATISs. **c**, Localization of unwound mRNA structure 3′ of activated ATISs. Enrichment of differential DMS-MaPseq counts. Negative values indicate unwound mRNA regions in wild-type *DED1* within 20 nt of high-stringency ATISs ($n = 274$), compared to all other near-cognate codons ($n = 60{,}666$; excluding ATIS). The shaded area marks a significant difference in secondary RNA structure between the regions downstream of an ATIS and downstream of any near-cognate codon ($P = 0.00004$, two-tailed t-test). The dashed line marks the 5′ nucleotide of the ATISs or the near-cognate codon.

between increased ribosome occupancy in the 5′ UTR and diminished translation of the main ORF, we performed polysome fractionation with northern blot analysis of individual mRNAs. The *PSA1* mRNA—the translation efficiency of which is markedly affected by Ded1p—showed a distinct shift from polysomes to monosomes in *ded1-95* compared to the wild type, upon temperature shift but not at 30 °C (Fig. 1c, d, Extended Data Fig. 3c). *TDH2* mRNA, which is largely unaffected by Ded1p, did not show a comparable shift (Fig. 1c, d). Collectively, these observations suggest that increased ribosome occupancy on 5′ UTRs correlates with binding of the mRNA to only a single ribosome. This notion is consistent with previous reports[12]. Ribosome profiling on only the 80S monosome fraction upon temperature shift also showed more footprints on 5′ UTRs in the *ded1-95* strain, compared to the wild type (Extended Data Fig. 3d), indicating that ribosome occupancy on 5′ UTRs broadly correlates with the binding of mRNAs to single ribosomes.

A large number of sites on 5′ UTRs with increased ribosome footprints in the *ded1-95* strain were enriched with near-cognate initiation codons (Fig. 1e–g), which differ from the canonical 5′-AUG-3′ initiation codon by a single nucleotide and can create alternative translation initiation sites (ATISs)[13]. Increased ribosome occupancy on AUG codons in 5′ UTRs was also seen in the *ded1-95* strain (Fig. 1f), but only a few of these sites exist in the yeast transcriptome, compared to nine different near-cognate initiation codons, which constitute roughly 14% of all codons[14]. Ribosomes can translate from the *ded1-95*-activated ATISs, as demonstrated by ribosome profiles on small ORFs that start at these ATISs and finish at the respective termination codons, by the lack of ribosome accumulation at ATISs when translation was not arrested and by the periodicity of ribosome footprints starting from ATISs (Extended Data Fig. 3e–l). Collectively, the data indicate that defective Ded1p leads to ATIS activation in 5′ UTRs, which decreases polysome formation on the main ORFs and thereby overall protein production. We conclude that Ded1p function suppresses the use of ATISs.

Although ATIS activation in *ded1-95* was extensive, only a subset of all near-cognate initiation codons was used. We detected no preferred length or register of the corresponding small ORFs relative to the main ORFs. However, in the ATISs, near-cognate codons from which translation initiation is most efficient were over-represented, whereas near-cognate codons from which translation initiation is least efficient were underrepresented[15,16] (Extended Data Fig. 4a). These observations show that ATIS activation is influenced by inherent codon preferences of the pre-initiation complex (PIC), although these preferences do not fully explain the ATIS activation pattern (Extended Data Fig. 4b–e).

To better understand this pattern, we examined whether remodelling of mRNA secondary structure by Ded1p is linked to ATIS activation. As an RNA helicase, Ded1p has been implicated in RNA structure remodelling[11,17], but it is not known which mRNA structures Ded1p alters in cells. To delineate the cellular mRNA structures that are remodelled by Ded1p, we used dimethyl sulfate (DMS) probing in vivo[18] and measured changes in mRNA structure in *ded1-95* and wild-type strains upon temperature shift (Extended Data Fig. 5a, b). Unwinding of mRNA structure by Ded1p was most pronounced in 5′ UTRs, compared to other mRNA regions (Fig. 2a, Extended Data Fig. 5c). Notably, *ded1-95* activated ATISs were generally located 5′ of unwound RNA regions (Fig. 2b, c). Even near-cognate codons for which translation initiation is least efficient were activated, if they were located 5′ of mRNA structure (Fig. 2b). Our observations link the inability of *ded1-95* to resolve mRNA structure to ATIS activation, suggesting that Ded1p suppresses ATIS activation by unwinding mRNA structure.

To investigate how Ded1p physically accomplishes this function, we determined which cellular RNAs bound to wild-type Ded1p using an high-throughput cross-linking-based approach (cross-linking-aided RNA affinity precipitation with sequencing (XL-RAP–seq)) and the individual-nucleotide resolution cross-linking and immunoprecipitation (iCLIP) technique to map Ded1p-binding sites on these RNAs[19] (Extended Data Fig. 6a–c). Ded1p cross-linked predominantly to mRNAs and ribosomal RNA (Extended Data Fig. 6b), especially to the 40S ribosomal subunit (Fig. 3a), which is part of the PIC that scans 5′ UTRs[20]. The most frequently cross-linked position maps to helix 16, located at the mRNA entry channel (Fig. 3a). Notable cross-linking was also observed at helix 26, which is located at the mRNA exit site, and in extension segment 6 (around nucleotide 720), which is located in the vicinity of the other cross-link sites on the solvent side of the 40S subunit (Fig. 3a). Ded1p binding to helices 16 and 26 is consistent with reported interactions between Ded1p and eIF3c and the eIF3b–eIF3g–eIF3i sub-complex, that binds near these sites[21,22] (Extended Data Fig. 7a, b). Human DDX3X also binds to helix 16[3].

Ded1p further cross-linked to virtually all expressed mRNAs, predominantly in 5′ UTRs (Fig. 3b, c). This cross-linking pattern is consistent with the physical contact of Ded1p to the PIC. Aside from a modest preference for A and U, no sequence motifs could be identified in the mRNA cross-linking sites (Extended Data Fig. 7c). However, peaks of Ded1p cross-linking on 5′ UTRs were frequently proximal to *ded1-95*-activated ATISs (Fig. 3c, d), and unwound mRNA structure was located 3′ of Ded1p cross-linking sites (Fig. 3c, f).

Collectively, the data link Ded1p binding to mRNA, unwound mRNA structure, ATIS location and binding of Ded1p to the PIC. This link is illustrated by a segment of the *PSA1* mRNA as an example (Fig. 4a). Ded1p binding is most pronounced 5′ of unwound RNA

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

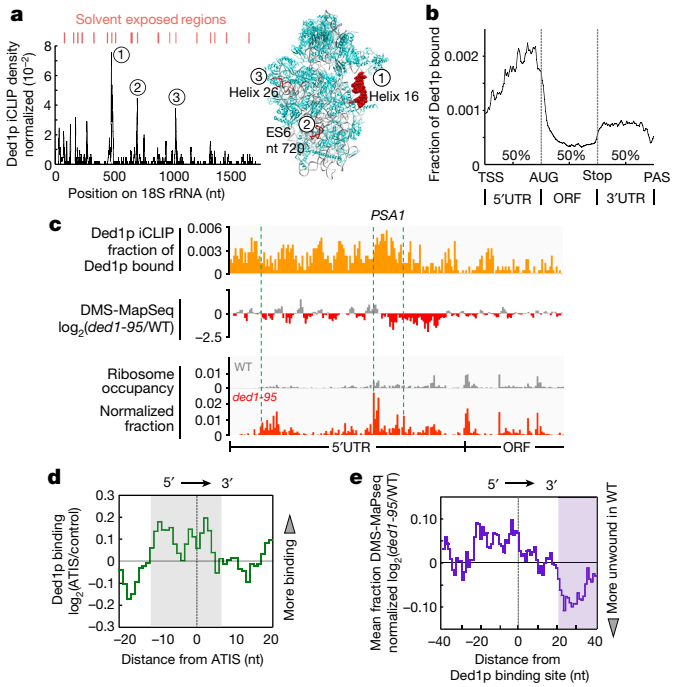


Fig. 3 | Ded1p cross-linking to the 40S ribosomal subunit and to mRNAs. a, Left, fraction of iCLIP reverse transcription stops on 18S rRNA. Moving average of ±2 nt, values represent the average from two independent experiments. Numbers denote predominant cross-linking sites. Right, the position of the three predominant Ded1p cross-linking sites (red) in the crystal structure of the 40S ribosomal subunit[30]. RNA, grey; ribosomal proteins, cyan; Ded1p cross-link sites, red. ES6, extension segment 6. **b**, Metagene profile of Ded1p association to mRNAs, calculated from two independent iCLIP experiments (moving average of ±1 nt). Stop, translation stop site; PAS, polyadenylation site. **c**, Ded1p cross-linking to the 5′ UTR of the *PSA1* mRNA. Top, fraction of reverse transcription stops per nucleotide, normalized to transcript length. For comparison, differential DMS-MapSeq (middle) and ribosome profiling (bottom) tracks of the 5′ UTR of *PSA1* mRNA for wild-type *DED1* and *ded1-95* are shown (5 min, 37 °C). Similar results were obtained in two independent experiments. **d**, Enrichment of Ded1p cross-linking within 20 nt of ATISs ($n = 274$) normalized to the background distribution of Ded1p binding (moving average of ±1 nt, reverse transcription stops normalized for each mRNA). The dashed line marks the ATIS position. The shaded area marks a significant difference in Ded1p binding between the regions in the vicinity of an ATIS and in the vicinity of a random position within the same 5′ UTR ($P = 0.013$, two-tailed *t*-test). **e**, Enrichment of differential DMS-MaPseq counts (Fig. 2c) within 40 nt of Ded1p binding sites ($n = 178$, high-stringency ATIS) on 5′ UTRs. The shaded area marks a significant difference in RNA structure between the regions downstream of a Ded1p binding site and downstream of a random position within the same 5′ UTR ($P = 0.008$, two-tailed *t*-test).

structure, indicating that Ded1p does not exclusively contact mRNA structure, but also regions that are 5′ of the structure. This finding is consistent with the notion that Ded1p functions in the context of the scanning PIC. The scanning process is slowed by RNA structure[20,23], and a slowed PIC conceivably permits Ded1p to survey the mRNA for structured regions that it then unwinds. Biochemical data show higher functional affinity of Ded1p for unstructured RNA, compared to structured RNA[8], rationalizing the contacts of Ded1p to unpaired mRNA that is 5′ of unwound mRNA structure, as the helicase travels in a 5′ to 3′ direction with the PIC.

Our data collectively indicate that failure of Ded1p to resolve mRNA structure leads to ATIS activation. To directly probe the link between mRNA unwinding and ATIS activation, we generated a *PSA1* mRNA with a mutation in an activated ATIS 5′ of unwound RNA structure. (Fig. 4b). The mutation markedly diminished the sensitivity to Ded1p-deficiency seen with the native *PSA1* mRNA (Fig. 4b, c). Alterations in the RNA structure 3′ of the ATIS also decreased sensitivity to Ded1p (Extended Data Fig. 8a, b). Identical observations were made for mutations in an ATIS and the corresponding RNA structure in the *ATP5* mRNA (Extended Data Fig. 8c–f). These results show that the effect of Ded1p on translation initiation depends not only on RNA unwinding, but also on proximal ATISs. Without a proximal ATIS, failure of Ded1p to unwind 5′ UTR structures does not abrogate scanning of the PIC and subsequent translation of the main ORF (Extended Data Fig. 9). This finding challenges the notion that cellular 5′ UTR structures alone are insurmountable hindrances for the scanning PIC.

Together, our results suggest the following function for Ded1p on 5′ UTRs (Fig. 4d). The enzyme associates with the PIC in the vicinity of the mRNA entry site of the small ribosomal subunit (Fig. 3a). This site is in close proximity to eIF4G and eIF4A (Extended Data Fig. 7b), both of which bind Ded1p with high affinity and might therefore be important for recruitment and function of Ded1p on the PIC[8,24,25]. The density of Ded1p cross-linking sites on 5′ UTRs increases with distance from the 5′ cap (Fig. 3b), suggesting gradual recruitment of Ded1p to the mRNA entry site during the scanning process. This notion is consistent with the reported increase of Ded1p function with greater distance from the 5′ cap and with 5′ UTR length[11]. The mRNA binding pattern of Ded1p further suggests that Ded1p is targeted to its sites of action through association with the scanning PIC. This is an effective way to deploy the enzyme exactly at sites at which it is needed, even though these sites lack common sequence or defined structure signatures. If Ded1p is missing or defective, mRNA structure persists, the PIC stalls and either dissociates from the mRNA, continues slowed scanning through the structure, or undergoes subunit joining and translation initiation if a near-cognate codon is present (Fig. 4d). Ribosomes initiating on an ATIS block subsequent scanning ribosomes from reaching the canonical initiation site, thereby decreasing translation efficiency for the main ORF (Fig. 1b). Unless an ATIS marks an N-terminal extension of the main ORF, PICs initiating at an ATIS are likely to be deterred from translating the main ORF. PICs encountering 5′ UTR structures without a proximal ATIS also interfere with scanning, but the kinetic pause introduced by PIC stalling, slowed scanning through the structure or a combination thereof is shorter than on an activated ATIS. Slowed PICs will eventually reach the main ORF (Extended Data Fig. 9), and therefore 5′ UTR structure alone affects main ORF translation less in isolation than it does in combination with proximal ATISs. Our model for Ded1p function does not preclude additional roles of the enzyme before the PIC scanning process[25]. However, the Ded1p function outlined above largely accounts for the observed Ded1p interactions with mRNA, and therefore, additional roles of Ded1p are probably restricted to transient Ded1p–mRNA interactions.

Finally, our data reveal a straightforward mechanism for activation of upstream ORFs. The mRNA structures in the 5′ UTRs represent a large set of riboswitches that are sensitive to Ded1p. Active Ded1p turns the switches off, suppresses ATIS activation and allows efficient translation of the main ORF. Inactivation of the helicase by post-translational modifications[1], by metabolites such as AMP[26], by decreased Ded1p levels or by sequestration of Ded1p in RNP granules[25,27] turns the switches on, activating the ATISs and thereby inhibiting translation from the corresponding main ORFs. Certain peptides that are translated from activated ATISs might also have direct biological functions[28], but the regulation described here appears to be independent of functional peptides.

This mechanism for activation of upstream ORFs is probably used in biological processes. This notion is supported by several lines of evidence. First, there is a marked increase in sequence conservation in the RNA regions around activated ATIS (Extended Data Fig. 10a). Second, the *ded1-95* activated ATIS in the *ALA1* transcript produces an N-terminal extension that targets Ala1p to the mitochondria[29] (Extended Data Fig. 10b). Third, during meiosis, ATIS activation[15] occurs in a pattern that is highly similar to the ATIS activation pattern seen with *ded1-95* upon temperature shift (Fig. 4e, f, Extended Data Fig. 10c). Notably, we find reduced Ded1p levels during meiosis (Fig. 4g, Extended Data Fig. 10d). This link between Ded1p levels and

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

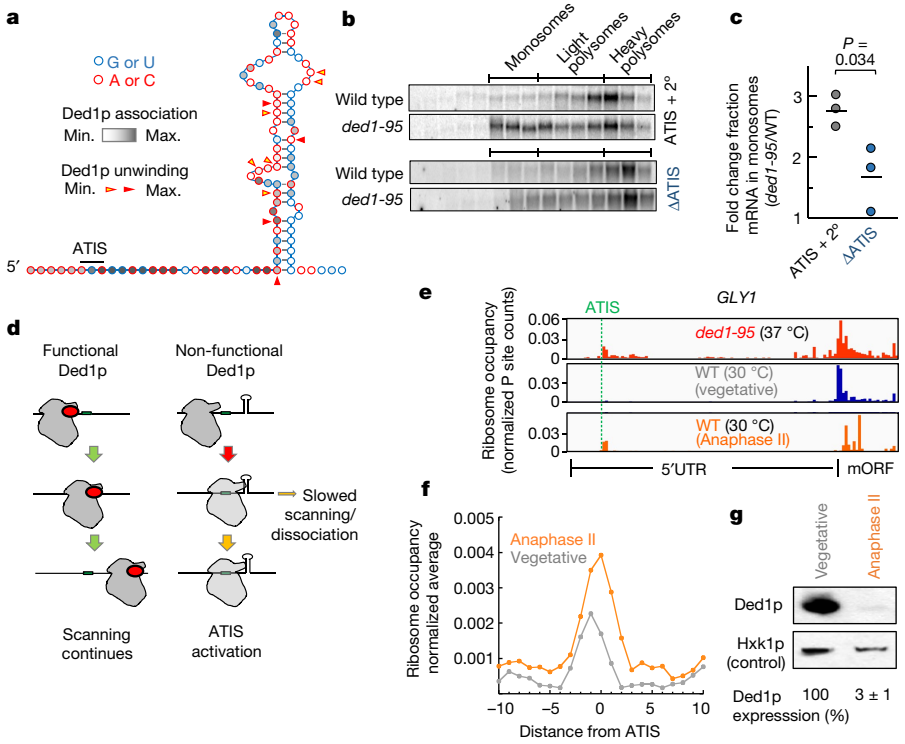


Fig. 4 | Ded1p function on 5′ UTRs. a, DMS-MapSeq constrained secondary structure model of a fragment of the *PSA1* mRNA 5′ UTR. The ATIS is marked by a line. Shading indicates Ded1p cross-linking (iCLIP). Triangles indicate unwinding (DMS-MapSeq) for each nucleotide ($\log_2$ ratio of normalized DMS-MapSeq counts of wild type/*ded1-95* in two categories). Yellow triangles, 0.35–0.7 (moderately unwound) and red triangles, >0.7 (strongly unwound). **b**, Representative RNA blots, after sucrose gradient centrifugation for wild-type *DED1* and *ded1-95*, (5 min, 37 °C) for constructs expressing *PSA1–FLAG* mRNA with wild-type 5′ UTR (ATIS and secondary structure (Δ2°)) or with a mutated ATIS (ΔATIS). Similar results were obtained in three independent experiments. **c**, Quantification of RNA blot experiments shown in **b**, indicating the fold change in the fraction of *PSA1* mRNA in monosomes in *ded1-95*, compared to the wild type upon temperature shift. Data are from three independent biological replicates, lines mark the mean. *P* value from a two-tailed *t*-test. **d**, Schematic of Ded1p function on 5′ UTRs. mRNA is depicted as a black line, mRNA structure as a hairpin, the PIC as a grey shape, Ded1p as a red oval, and the near-cognate codon as a green rectangle. **e**, Ribosome occupancy tracks (5 min, 37 °C) of *ded1-95* and wild-type *DED1* (vegetative control and anaphase II) for *GLY1* mRNA. ATISs are marked by dashed lines. Similar results were obtained in two (vegetative control) and four (anaphase II) independent experiments. mORF, main ORF. **f**, Mean ribosome occupancy 10 nt 3′ and 5′ of a high-confidence ATIS on 5′ UTRs (moving average ± 1nt) for wild-type *DED1* (vegetative control and anaphase II). **g**, Representative western blot of Ded1p and Hxk1p (loading control) in vegetative cells and cells in anaphase II. Numbers indicate the relative expression level of Ded1p from four independent experiments with s.d.

the activation of ATISs proximal to 5′ UTR structures during meiosis suggests a role for the levels of Ded1p in this process. Collectively, our observations show that the regulatory program linking Ded1p to mRNA structure and ATIS activation is used in a physiological cellular process. The results indicate that intricate translation control and activation of upstream ORFs can be based on simple, ubiquitous elements: a helicase, mRNA structure and near-cognate initiation codons.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0258-0.

Received: 30 January 2017; Accepted: 8 May 2018;
Published online: 27 June 2018

1. Sharma, D. & Jankowsky, E. The Ded1/DDX3 subfamily of DEAD-box RNA helicases. *Crit. Rev. Biochem. Mol. Biol.* **49**, 343–360 (2014).
2. Bol, G. M., Xie, M. & Raman, V. DDX3, a potential target for cancer treatment. *Mol. Cancer* **14**, 188 (2015).
3. Oh, S. et al. Medulloblastoma-associated DDX3 variant selectively alters the translational response to stress. *Oncotarget* **7**, 28169–28182 (2016).
4. Pugh, T. J. et al. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* **488**, 106–110 (2012).
5. Snijders Blok, L. et al. Mutations in DDX3X are a common cause of unexplained intellectual disability with gender-specific effects on Wnt signaling. *Am. J. Hum. Genet.* **97**, 343–352 (2015).
6. Valiente-Echeverría, F., Hermoso, M. A. & Soto-Rifo, R. RNA helicase DDX3: at the crossroad of viral replication and antiviral immunity. *Rev. Med. Virol.* **25**, 286–299 (2015).
7. Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protocols* **7**, 1534–1550 (2012).
8. Putnam, A. A. et al. Division of labor in an oligomer of the DEAD-box RNA helicase Ded1p. *Mol. Cell* **59**, 541–552 (2015).
9. Burckin, T. et al. Exploring functional relationships between components of the gene expression machinery. *Nat. Struct. Mol. Biol.* **12**, 175–182 (2005).
10. Chuang, R. Y., Weaver, P. L., Liu, Z. & Chang, T. H. Requirement of the DEAD-box protein Ded1p for messenger RNA translation. *Science* **275**, 1468–1471 (1997).
11. Sen, N. D., Zhou, F., Ingolia, N. T. & Hinnebusch, A. G. Genome-wide analysis of translational efficiency reveals distinct but overlapping functions of yeast DEAD-box RNA helicases Ded1 and eIF4A. *Genome Res.* **25**, 1196–1205 (2015).
12. Heyer, E. E. & Moore, M. J. Redefining the translational status of 80S monosomes. *Cell* **164**, 757–769 (2016).
13. Hinnebusch, A. G., Ivanov, I. P. & Sonenberg, N. Translational control by 5′-untranslated regions of eukaryotic mRNAs. *Science* **352**, 1413–1416 (2016).
14. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
15. Brar, G. A. et al. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **335**, 552–557 (2012).
16. Kolitz, S. E., Takacs, J. E. & Lorsch, J. R. Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation. *RNA* **15**, 138–152 (2008).
17. Berthelot, K., Muldoon, M., Rajkowitsch, L., Hughes, J. & McCarthy, J. E. Dynamics and processivity of 40S ribosome scanning on mRNA in yeast. *Mol. Microbiol.* **51**, 987–1001 (2004).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

18. Zubradt, M. et al. DMS-MaPseq for genome-wide or targeted RNA structure probing *in vivo*. *Nat. Methods* **14**, 75–82 (2017).
19. Huppertz, I. et al. iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* **65**, 274–287 (2014).
20. Hinnebusch, A. G. The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.* **83**, 779–812 (2014).
21. Gavin, A. C. et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
22. Krogan, N. J. et al. High-definition macromolecular composition of yeast RNA-processing complexes. *Mol. Cell* **13**, 225–239 (2004).
23. Kozak, M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl Acad. Sci. USA* **87**, 8301–8305 (1990).
24. Gao, Z. et al. Coupling between the DEAD-box RNA helicases Ded1p and eIF4A. *eLife* **5**, e16408 (2016).
25. Hilliker, A., Gao, Z., Jankowsky, E. & Parker, R. The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex. *Mol. Cell* **43**, 962–972 (2011).
26. Putnam, A. A. & Jankowsky, E. AMP sensing by DEAD-box RNA helicases. *J. Mol. Biol.* **425**, 3839–3845 (2013).
27. Jain, S. et al. ATPase-modulated stress granules contain a diverse proteome and substructure. *Cell* **164**, 487–498 (2016).
28. Andrews, S. J. & Rothnagel, J. A. Emerging evidence for functional peptides encoded by short open reading frames. *Nat. Rev. Genet.* **15**, 193–204 (2014).
29. Tang, H. L. et al. Translation of a yeast mitochondrial tRNA synthetase initiated at redundant non-AUG codons. *J. Biol. Chem.* **279**, 49656–49663 (2004).
30. Aylett, C. H., Boehringer, D., Erzberger, J. P., Schaefer, T. & Ban, N. Structure of a yeast 40S–eIF1–eIF1A–eIF3–eIF3j initiation complex. *Nat. Struct. Mol. Biol.* **22**, 269–271 (2015).

Acknowledgements This study was supported by the National Institutes of Health (NIH) (GM118088 to E.J., GM107331 to D.D.L.) and by a postdoctoral fellowship from the German Research Council (GU 1146/1-1 to U.-P.G.). D.P.B. and J.S.W. are investigators of the Howard Hughes Medical Institute. We thank M. Adams for help with the initial data analysis; N. Al-Huseini and J. Coller for discussion and materials; A. Tambe, M. Hannigan, and R. Backofen for advice on bioinformatic data analysis; B. Klaus for critical advice on statistical analysis and M. Hentze for discussion.

Reviewer information Nature thanks P. Todd and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions U.-P.G.: conceptualization, material generation, ribosome profiling, XL-RAP–seq, iCLIP, molecular biology techniques, bioinformatic analysis, statistical analysis and manuscript writing; D.E.W.: conceptualization and iCLIP; M.M.Z.: conceptualization, DMS-MaPseq and bioinformatic analysis; F.A.T.: iCLIP, bioinformatic analysis, western blots; B.N.S.: material generation and northern blots; L.L.Z. and D.D.L.: iCLIP; G.A.B.: material generation and meiosis strains; D.P.B.: conceptualization and study supervision; J.S.W.: conceptualization and study supervision.; E.J.: conceptualization, study supervision and manuscript writing. All authors commented on and edited the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0258-0.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0258-0.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to E.J.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Yeast strains, plasmids and oligonucleotides. Yeast strains used in this study are listed in Supplementary Table 1. Strains were grown at 30 °C unless stated otherwise. Primers, northern blot probes and other DNA oligonucleotides are listed in Supplementary Table 2, RNA oligonucleotides are listed in Supplementary Table 3, and DNA plasmids are listed in Supplementary Table 4.

Generation of a yeast strain expressing Ded1p–HTBH. Construction of plasmid pEJ21 containing the N-terminally haemagglutinin (HA)-tagged Ded1p has previously been described[24]. pEJ21was then used to generate the plasmid pEJ5. The HA tag was replaced by a sequence containing a HpaI and a SphI site (amplification with primers X1 and X2), generating pEJ1. The His_6-TEV-Biotin (HTB) tag was amplified from pFA6-HTB-kanMX6 plasmid (gift from P. Kaiser) with primers X3 and X4.The resulting PCR product was cloned into pEJ1 via its HpaI and SphI sites yielding pEJ2. A second His_6 tag was introduced by site-directed mutagenesis with primers X5 and X6 generating pEJ3. The C-terminal His_6-TEV-Biotin-His_6 (HTBH) tag was introduced into pEJ4[24] by amplification of pEJ3 with primers X8 and X9 and subcloning with PflMI and SpeI into pEJ4, yielding pEJ5. pEJ5 was linearized and used to transform BY4741 by standard lithium acetate transformation yielding yeast strain yEJ2.

Generation of a yeast strain expressing Ded1p–His_6–$FLAG_3$. Yeast strain yDPB740, containing a C-terminal His_6–$FLAG_3$ tag on the endogenous *DED1* allele, was generated from BY4742 using standard methods. In brief, a homologous recombination template was designed comprising the 40 nucleotides upstream and downstream of the DED1 stop codon flanking the His_6FLAG_3 tag (with stop codon) and kanMX6 drug resistance cassette. This template was generated by amplifying from pFA6a–6×His–3×FLAG–kanMX6 plasmid with primers DW1 and DW2. PCR product was used to transform BY4742 by standard lithium acetate transformation yielding yeast strain yDPB740.

Generation of a yeast strain expressing wild-type, ΔATIS and secondary structure mutants of *PSA1* and *ATP5* mRNAs. FLAG-tagged PSA1 and ATP5 strains were generated from the respective cDNAs using standard methods as described above (pEJ14, pEJ15 and pEJ18, pEJ19, respectively). The FLAG-tag was appended at the 3′ terminus of the *PSA1* and *ATP5* ORF, respectively. For *PSA1*, mutations in the ATIS in the 5′ UTR (PSA1-ΔATIS, pEJ16, Fig. 4b, c) contained the following changes: (c.-58A>C (c.-58 indicates the 58th nucleotide 3′ of the A of the AUG), c.-56A>C, c.-52T>A). Mutations in the 5′ UTR mRNA structure (PSA1-Δ2°, pEJ17, Extended Data Fig. 8a, b), contained the following changes: (c.-39_- 38AG>TC, c.-36_-35TA>AT, c.-32A>T,c.-24_-22AAA>TCT, c.-19_-18AA>CT). For ATP5, mutations in the ATIS in the 5′ UTR (ATP5-ΔATIS, pEJ20, Extended Data Fig. 8e, f) contained the following changes: (c.-126A>C, c.-120_119TT>CC, c.-109A>C). Mutations in the 5′ UTR mRNA structure (ATP5-Δ2°, pEJ22, Extended Data Fig. 8e, f), contained the following changes: (c.-102C>A, c.-99C>A, c.-96G>A, c.-83G>A, c.-81C>A, c.-76G>A, c.-74G>A).

Polysome analysis. Polysome analysis using 20U (A_{260}) lysate was performed as described[31], with a lower final concentration of cycloheximide (50 μg ml^{-1}). In brief, after centrifugation through a 15–45% (w/v) sucrose gradient, sixteen fractions were collected at a pump speed of $S = 0.9$ ml min^{-1}. RNA in each fraction was precipitated by adding two volumes of ice-cold ethanol and incubating overnight at −80 °C. RNA was extracted with phenol–chloroform according to standard protocols. Samples were applied to a 1.4% agarose gel containing 6% (v/v) formaldehyde, and electrophoresis was performed as described[31]. RNA was visualized with ethidium bromide. The amount of 18S rRNA in each fraction of the gradient was quantified with ImageQuant 5.2 software (Molecular Dynamics).

Northern blot analysis. For ribosome association of individual mRNAs, gel electrophoresis was performed after polysome analysis and fractionation as described above. RNA was subsequently transferred to nitrocellulose membranes (AmershamHybond–N, GE Healthcare) and further processed as described[31]. DNA oligonucleotides (Supplementary Table 2; X85, X96, X105, X114) were radiolabelled with PNK according to standard procedures. Probes were incubated with the membranes in hybridization buffer (6× saline sodium citrate (SSC), 0.1% SDS, 10× Denhardt's reagent) overnight at 42 °C. Membranes were subsequently washed three times with wash buffer (6× SSC, 0.1% SDS). Probe signals were visualized using a Molecular Dynamics Phosphorimager (GE Healthcare) and quantified with ImageQuant 5.2 software (Molecular Dynamics). Normalized signal intensities were compiled for fractions corresponding to monosomes, light and heavy polysomes and averaged from at least three biological replicates.

Western blot analysis. Lysates from yeast strain A14201 at vegetative phase and stage 11 (ndt80 release time course) were prepared as described[15]. After loading equal amounts of protein on a 10% NEXT gel, denaturing gel electrophoresis and transfer to a PVDF membrane, western blotting was performed with anti-Ded1p (rabbit; 1:5,000) and anti-hexokinase (rabbit;1:10,000; US Biological) antibodies. Chemiluminescence was quantified by Imagequant software. Hexokinase served as a loading control for normalization.

Ribosome profiling. Yeast cultures (500 ml) were grown at 30 °C in rich medium to mid-log phase ($OD_{600\,nm}$ of approximately 0.4) and divided into two equal volumes. Cycloheximide (final concentration, 50 μg ml^{-1}) was added to one sample. Cells were rapidly collected by centrifugation at 4,000*g* for 2 min, and snap-frozen on dry ice. One volume of pre-warmed medium (44 °C) was added to the remaining sample, resulting in a temperature of 37 °C for the entire volume. The temperature of 37 °C was verified. The entire sample was immediately moved to a shaking incubator at 37 °C. Five minutes after temperature shift, the yeast culture was treated with cycloheximide and cells were collected as mentioned above. For run-off experiments, yeast cells were collected in the absence of cycloheximide (Extended Data Fig. 3e).

Cell lysis, RNase I treatment (Ambion) and sucrose gradient centrifugation was performed as described[32] for 25 units A_{260} per sample. In addition, lysates (approximately 150 μl) were treated with 3.25 μl Turbo DNase I for 1 h at 25 °C. Purification and processing of ribosome-protected fragments were carried out as described[32], except that rRNA depletion with the RiboZero kit (Illumina) was omitted. Depletion of rRNA was performed at the level of circularized cDNA, as described[7] (1 μl of a 5 μM mix of biotinylated DNA oligonucleotides, Supplementary Table 2 (X66–X80) and MyOne Streptavidin C1 DynaBeads (Invitrogen)). PCR amplification and sequencing was performed as described[7].

Monosome-protected fragments were isolated as described[12]. In brief, sucrose gradient centrifugation of lysates was performed and fractions corresponding to monosomes were pooled and treated with 1/10 U RNase I per unit $A_{260\,nm}$ lysate and 0.4U Turbo DNase I per unit $A_{260\,nm}$ lysate for 1 h at 25 °C. Reactions were stopped by phenol–chloroform extraction of RNA. Monosome-protected fragments were processed as described above for ribosome-protected fragments.

The fragmented mRNA control libraries were generated as described[32]. Sizing, concentration and quality of each DNA library was assessed with the High Sensitivity DNA kit on an Agilent2100 Bioanalyzer system. Up to eight DNA libraries were pooled before performing 50 bp single end read sequencing on an Illumina HiSeq2500 V2 in rapid run mode.

Processing of the ribosome profiling data was performed as described[32]. In brief, adaptor sequences and ribosomal reads were removed. Remaining reads were mapped to the sacCer3 genome with the TopHat software (parameters set as: --no-novel-juncs -N 2 --read-edit-dist 2 --max-insertion-length 3 --max-deletion-length 3 -g 2 (https://www.yeastgenome.org)). All other parameters were kept at default settings[33]. The abundance of mRNAs in ribosome or monosome-protected fragments as well as in the fragmented RNA control libraries were determined using Cufflinks software[33]. These values were used to calculate translational efficiencies as described[14]. For the calculation of $\log_2\Delta TE$ values we also included a constant factor reflecting the change in the overall size of the mRNA pool, derived from the spike-in of RNA controls (Supplementary Table 3).

P-sites in ribosome-protected fragments (RPFs) were determined using the 13th position from the 5′ end of reads with 28 or 29 nt[14]. The fraction of ribosomes on 5′ UTRs was calculated for each mRNA by counting all RPFs on the 5′ UTR (excluding positions −3 to −1), divided by the number of all RPFs mapped to the entire mRNA. mRNAs with 5′ UTRs containing fewer than 10 nt were excluded from the analysis. The centre of ribosome density (CRD) was calculated as described[34]. The shift in the CRD (ΔCRD) in *ded1-95* compared to wild-type *DED1* upon temperature shift was defined relative to the entire length of the mRNA according to:

$$\Delta CRD = ((CRD\ ded1{-}95) - (CRD\ \text{wild-type}\ DED1))/\text{mRNA length}$$

A negative ΔCRD value marks increased ribosome accumulation in the 5′ UTR in *ded1-95*.

ATIS were identified according to a previously described algorithm[15]. In brief, a position is considered an ATIS, (i) if minimal ribosome count value (±1 nt of the nucleotide under consideration) is greater than nine (high-stringency ATIS) or four (medium-stringency ATIS) in all replicates; (ii) if the ratio of ribosome occupancy between two neighbouring nucleotides 5′ to 3′ ($position^{n-1}$/$position^{n}$) is greater than or equal to three (high-stringency ATIS) in all replicates, or greater or equal of three in one and greater than or equal to 1.75 in the other replicate (medium-stringency ATIS); and (iii) if the normalized ribosome count in *ded1-95* cells 5′ after temperature shift to 37 °C is 1.5-fold higher that in wild-type *DED1* in all replicates. This algorithm identified 396 high-stringency ATIS and 2,126 medium-stringency ATIS. Near-cognate codons were identified in 259 high-stringency ATIS (65%) and 1,382 medium-stringency ATIS (65%) within a moving window of ±1 nt. Canonical AUG initiation codons were found in 4% high-stringency ATIS, and in 3% medium-stringency ATIS. As a control set, we collected all near-cognate codons on 5′ UTRs of mRNA genes with a 5′ UTR length between 20 and 500 nt. After removal of near-cognate codons in medium-stringency ATIS, we identified 60,666 near-cognate codons.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

XL-RAP–seq. Yeast cells containing HTBH-tagged *DED1* were grown in rich medium to an $OD_{600\ nm}$ of 1.0–1.5, collected by brief centrifugation at 4,000*g*, re-suspended in ice-cold water or remaining YPD medium, transferred to a Petri dish, and subjected to UV-light in a Stratalinker (600 mJ cm^{-2}, 254 nm) on ice. Cells were washed in ice-cold water, sedimented by centrifugation for 5 min at 5,250*g*, frozen on dry ice and stored at −80 °C.

Frozen cells were lysed in QIA-1M buffer (100 mM NaH_2PO_4 pH 8, 10 mM Tris, 1M NaCl, 8 M Urea, 10 mM imidazole, 0.5% (w/v) IGEPAL, 2.5 mM β -mercaptoethanol, 1 mM PMSF, protease inhibitor cocktail (Roche)) with glass beads six times for 30 s in a Beadbeater system (Biospec products). Glass beads were removed, and lysates were centrifuged at 5,250*g* for 30 min. Cleared lysates were incubated with Ni^{2+}–Agarose (40 µl slurry per g dry pellet weight, pre- equilibrated in buffer QIA-1M; Qiagen) overnight at 4 °C. Ni^{2+}-beads were washed in 25 ml of wash buffer 1 (0.3 M NaCl, 10 mM Tris, 100 mM NaH_2PO_4, 8 M Urea, 10 mM imidazole) and sample was eluted with 10 ml elution buffer1 (0.3 M NaCl, 100 mM Tris, 50 mM NaH_2PO_4, 8 M Urea, 500 mM imidazole, 10% (v/v) glycerol). Eluates were then incubated with12.5 µl equilibrated streptavidin-conjugated agarose resin (Pierce Technologies) per g dry pellet weight overnight at 4 °C. Streptavidin beads were washed with 12.5 ml wash buffer 2 (0.3 M NaCl, 100 mM Tris, 8 M Urea, 0.5 mM EDTA)—containing 2% SDS, with wash buffer 2 (12.5 ml) without SDS, and with 1× TEV buffer. Beads were next incubated with 50 U AcTEV (Invitrogen) for 2 h at 4 °C. The sample was eluted with 2 × 0.9 ml TEV elution buffer (300 mM NaOAc pH 6, 8 M Urea, 0.5% NP-40). Eluates were incubated with 175 µl Proteinase K (4 mg ml^{-1}; Roche), and the reaction was stopped by standard phenol–chloroform extraction. Released RNA was precipitated by ethanol precipitation overnight. RNA was re-suspended in 1× Turbo DNase buffer and incubated with 0.4 µl TurboDNase (Ambion) in a final volume of 20 µl for 20 min at 37 °C. Turbo DNase was inactivated according to the manufacturer's instruction and the RNA was precipitated with ethanol.

cDNA was generated from the RNA sample using the 'template switch' activity of M-MLV reverse transcriptase[35]. Reactions were performed in 20 µl according to the instructions of the manufacturer mix (Invitrogen) with 0.5 µl Superscript II and 0.05 µM final concentration of tailed random hexamer primer (X13, Supplementary Table 2). Reaction conditions were as follows: 20 °C, 10 min; 37 °C, 10 min; 42 °C, 45 min. Next, 0.25 µM of primers (X14–X17, Supplementary Table 2) containing a 7-nt index and three guanosine ribonucleotides at their 3′ end were added to the reaction mix and reactions were performed at 42 °C for 30 min. Four microlitres of the resultant cDNA were used as template for amplification with primers X18 and X19 (Supplementary Table 2) and Advantage 2 polymerase mix (Clontech) for 30 cycles, according to the manufacturer's instructions. PCR products were precipitated and washed with 75% ethanol. DNA libraries were sequenced as 36-bp single-end reads on an Illumina Genome Analyzer. Reads were mapped to sacCer2 genome with Bowtie software with default settings.

Reads were excluded from further analysis if their location was outside of the boundaries of an mRNA or other transcribed regions of the genome as previously defined[36]. Genes were only considered to be bound by Ded1p, if more than 10 FPKM mapped to the respective mRNA gene.

iCLIP. iCLIP experiments were performed independently with two different approaches, (i) using Ded1p–HTBH and (ii) Ded1p–His_6–$FLAG_3$, both on the endogenous *DED1* allele. For iCLIP with Ded1p–HTBH, cell growth, cross-linking and tandem affinity purification on Ni^{2+}-agarose and streptavidin beads was performed as described above for the XL-RAP–seq procedure. Streptavidin beads were washed twice in 1× PNK buffer and split into two samples (80%, L; 20%, H). 200 ng RNase I was added per gram dry pellet weight to the L-sample and 0.1 ng RNase I to the H-sample. The samples were incubated for 5 min at 37 °C on a rotator. Resins were subsequently washed with ∼1.5 ml wash buffer 2 (2% (w/v) SDS), and then with 1.5 ml wash buffer 2 without SDS, and finally with 1.5 ml 1× PNK buffer. The supernatant was removed and beads were re-suspended in 67 µl RNase-free water, 3 µl alkaline phosphatase (NEB) and 2 µl RNasin (Roche), and incubated for 20 min at 37 °C. Beads were washed twice with 1.5 ml 1× PNK buffer.

3′ ligation was performed by re-suspending the resin in 32 µl RNase-free water with 8 µl of 20 µM RL3 RNA Linker (Supplementary Table 3). Reactions were performed overnight at 4 °C in 40 µl, containing 22 µl RNase-free water, 8 µl 10× T4 RNA ligase buffer, 8 µl BSA (0.2 µg $µl^{-1}$), 3 µl T4 RNA ligase (all NEB).

The resin was washed with 1.5 ml 1× TEV-salt buffer (50 mM Tris pH 7.5, 300 mM NaCl, 0.5 mM EDTA, 1 mM DTT) and twice with 1.5 ml 1× PNK buffer. Beads were re-suspended in 64 µl RNase-free water, 8 µl 10× PNK Buffer, 4 µl ^{32}P-γ-ATP, 4 µl T4 PNK (all NEB). Reactions were performed for 50 min in a shaking thermoblock at 37 °C. Beads were washed twice with 1.5 ml TEV salt buffer and twice with 1.5 ml TEV-elution buffer. Beads were subsequently mixed with SDS loading dye and subjected to PAGE on a 10% NEXT gel (Amresco) according to the manufacturer's conditions. Gels were subsequently blotted with nitrocellulose membranes (Amersham Protran, GE Healthcare) and exposed to X-ray film. RNA was then liberated by Proteinase K treatment as described[37]. The purified RNAs were re-suspended in 89 µl RNase- free water, 11 µl 10× DNase I Buffer, 5 µl RNasin, 5 µl RQ1 DNase (all Promega) and incubated for 20 min at 37 °C. The reaction was stopped by standard phenol–chloroform extraction and RNA was ethanol-precipitated overnight.

The DNase-treated RNA was re-suspended in 1.4 µl RNase-free water, 0.2 µl 10 mM dNTPs, 1 µl 20 nM reverse transcription primer X97 (Supplementary Table 2), and incubated for 5 min at 65 °C. Next, 0.8 µl 5× first-strand buffer, 0.2 µl 1M DTT, 0.2 µl RNase Inhibitor, 0.2 µl Superscript III (all Invitrogen) were added and incubated for 30 min at 50 °C. RNA was degraded by alkaline hydrolysis (after addition of 0.5 µl of 1N NaOH and incubation at 98 °C for 15 min). After addition of loading buffer, cDNA was applied to 10% denaturing PAGE and staining with SYBR Gold. Fragments of 100–125 nt were cut from the gel. The gel slices were crushed and cDNA was recovered by incubation in 500 µl diffusion buffer (20 mM Tris-HCl pH 7.5, 250 mM NaOAc, 1 mM EDTA, 0.25% (w/v) SDS) overnight at 4 °C with subsequent ethanol precipitation.

The cDNA was suspended in 15 µl RNase-free water and circularized with CircLigase I (Epicentre) according to the manufacturer's instructions. The circularized cDNA was used for amplification with Phusion polymerase (NEB) and primers X98 and X99 (Supplementary Information 2). PCR settings were: 30 s at 98 °C and then 24 PCR cycles (10 s at 98 °C, 30 s at 58 °C, 30 s at 72 °C). PCR products were applied to 10% non-denaturing PAGE and visualized by SYBR Gold. Products with 75–90bp were extracted from cut gel slices, and ethanol precipitated as described above. PCR products were amplified with Phusion polymerase and primers X100 and X101 for five cycles using the same PCR settings as above. PCR products were then separated on a 2% agarose gel, cut out and subjected to Illumina sequencing using primer X102.

For iCLIP with Ded1p–His_6–$FLAG_3$, cells were grown in SD–Trp medium to $OD_{600\ nm}$ = 0.5–0.6. The culture was subsequently transferred to a 245 mm × 245 mm × 25 mm square Petri dish. UV cross-linking was performed in a Stratalinker 2400 (150 mJ cm^{-2}, 254 nm) at room temperature. Cells were collected by centrifugation for 5 min at 2000*g*, washed twice in ice-cold PBS, frozen in liquid nitrogen, and stored at −80 °C.

Frozen cells were lysed in CLIP lysis buffer (50 mM Tris-HCl, pH 7.8, 300 mM NaCl, 1% Triton X-100, 1 mM PMSF, protease inhibitor cocktail (Roche)) with glass beads six times for 1 min in a Disruptor Genie system (Scientific Industries). Lysates were centrifuged after removal of the glass beads at 10,000*g*, twice for 5 min. Cleared lysates (∼26.5 A_{260} units) were incubated with anti-FLAG M2 Magnetic Beads (20 µl slurry pre-equilibrated in CLIP Lysis Buffer; Sigma) in a total volume of 1 ml overnight at 4 °C. Beads were washed twice in 1 ml FLAG Wash Buffer (50 mM Tris-HCl pH 7.8, 1 M NaCl, 0.1% NP-40) and twice in 1 ml FLAG Elution Buffer (50 mM Tris-HCl pH7.8, 150 mM NaCl, 0.1% NP-40). Proteins were eluted twice in 95 µl FLAG Elution Buffer containing 150 ng µl 3× FLAG tag peptide (Sigma). Pooled eluates were incubated with 10 µl RNase I (Ambion), diluted 1:500,000 in FLAG elution buffer for 15 min at room temperature. Reactions were quenched with 960 µl 8 M guanidine–HCl, 90 µl Dilution Buffer (600 mM Tris-HCl pH 7.8, 3.93 M NaCl), 6.4 µl 2 M imidazole, 10.8 µl 10% NP-40and 12.8 µl 500 mM β-mercaptoethanol. RNase-treated eluates were incubated further with Ni-NTA magnetic agarose beads (50 µl slurry pre-equilibrated in Ni-NTA binding buffer (50mM Tris-HCl pH 7.8, 300 mM NaCl, 10 mM imidazole, 6 M guanidine-HCl, 0.1% NP-40, 5 mM β-mercaptoethanol); Qiagen) overnight at 4 °C.

Ni^{2+}-beads were washed twice in 1 ml CLIP wash buffer I (50 mM Tris-HCl pH 7.8, 500 mM NaCl, 10 mM imidazole, 6 M guanidine-HCl, 0.1% NP-40, 5 mM β-mercaptoethanol) and three times in 1 ml 1× PNK Buffer (50 mM Tris-HCl pH 7.8, 10 mM $MgCl_2$, 0.5% NP-40, 10 mM β-mercaptoethanol). Beads were subsequently incubated with 30 µl dephosphorylation mix (50 mM Tris-HCl pH 7.8, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 3 M BU TSAP (Promega), 30 U SUPERase-In (Ambion)) for 30 min at 37 °C in a thermomixer at 1000 rpm. Reactions were terminated by adding 1 ml CLIP wash buffer I, and beads were washed three times in 1 ml 1× PNK buffer, re-suspended in 30 µl ligation mix (50 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 10% PEG8000 (NEB), 10% DMSO, 2 µM 3′ adenylated adaptor X103, 30 U T4 RNA Ligase 1 (NEB), 30 U SUPERase-In (Ambion)) and incubated for 3 h at 22 °C in a thermomixer at 1000 rpm. Reactions were terminated by adding 1 ml CLIP wash buffer I. Beads were washed three times in 1 ml 1× PNK Buffer, re-suspended in 30 µl Kinase Mix (50 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 30 U PNK (NEB), 5 µCi γ- ^{32}P-ATP, 30 U SUPERase-In (Ambion)) and incubated for 30 min at 37 °C in a thermomixer (1000 rpm). Reactions were terminated by adding 1 ml CLIP Wash Buffer I. Beads were washed three times in 1 ml CLIP Wash Buffer I, and three times in 1 ml CLIP wash buffer II (50 mM Tris-HCl, pH 7.8, 50 mM NaCl, 10 mM imidazole, 0.1% NP-40, 5 mM β-mercaptoethanol).

Proteins were eluted with 35 µl CLIP wash buffer II containing 200 mM imidazole three times for 5 min at 22 °C in a thermomixer (1000 rpm).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Pooled eluates were digested with Proteinase K (Invitrogen) at 50 °C for 1 h. The reaction was stopped by standard phenol–chloroform extraction. Released RNA was precipitated by ethanol precipitation overnight in the presence of 1 μl GlycoBlue (Ambion) and 1 pmol reverse- transcription DNA primer X104. RNA was converted to first-strand cDNA in a 10 μl standard reaction mix with 0.5 μl Superscript III (Invitrogen) and the co-precipitated DNA primer. The reaction conditions were as follows: 25 °C for 5 min, 42 °C for 20 min, 50 °C for 40 min. RNA was degraded with 1.67 μl of 1 M NaOH for 10 min at 90 °C. The cDNA was ethanol precipitated with GlycoBlue for >2 h. cDNA fragments of 120–200 nt were gel purified on a urea-denaturing 6% acrylamide gel and ethanol precipitated with GlycoBlue overnight. cDNA was circularized in a 10 μl reaction mix using 0.5 μl CircLigase I (Epicentre) in the presence of 1 M betaine at 60 °C for 1h. The reaction was then supplemented with additional 0.5 μl CircLigase I and incubated at 60 °C for 1 h. The enzyme was inactivated at 80 °C for 10 min. PCR and formamide-gel purification of PCR products were performed as described[38], using 20 cycles of PCR and isolating 120–200 nt ssDNA fragments.

Sequencing reads were processed as previously described[37,39]. In brief, after the trimming of adaptor sequences, reads were mapped to sacCer3 with Bowtie2 or TopHat software with similar settings used for the ribosome profiling data, outlined above. Identical reads were subsequently collapsed and duplications removed. The cross-link position of Ded1p to RNA was defined as 1 nt 5′ of the 5′ mapped nucleotide of a sequencing read[37].

DMS-MaPseq. Yeast strains (wild-type DED1 and *ded1-95*) were grown in YPD at 30 °C. Overnight cultures were diluted to $OD_{600\,nm}$ of ~ 0.09 and grown to an $OD_{600\,nm} = 0.6$. An equal volume of 44 °C YPD medium was added to achieve an immediate temperature shift to 37 °C, as outlined for the ribosome profiling experiments. Cultures were incubated in a 37 °C water bath for 3 min. At this time, DMS (Sigma) was added to a 5% (v/v) final concentration and incubation was continued with stirring for 3 min. DMS was quenched by adding 30 ml of ice-cold stop solution (30% β-mercaptoethanol, 50% (v/v) isoamyl alcohol). Cells were quickly transferred to ice, collected by centrifugation at 3,500*g* at 4 °C for 4 min, and washed with 10 ml 30% β- mercaptoethanol solution. Cells were re-suspended in 0.6 ml RNA lysis buffer (6 mM EDTA, 45 mM NaOAc, pH 5.5). Total RNA was purified with hot acid phenol (Ambion) and ethanol precipitation. Sequencing libraries were prepared as previously described[17].

Raw fastq files were stripped of linker sequences and filtered for overall quality using the FASTX-Toolkit Clipper and Quality Filter functions (http://hannonlab.cshl.edu/fastx_toolkit/), respectively, requiring that 80% of sequenced bases have a quality score >25. Reads were aligned against the yeast genome (sacCer3) using Tophat v2.1.0 with Bowtie2 with the following settings for a 50 bp sequencing run: --no-novel-juncs -N 5 --read-gap-length 7 --read-edit-dist 7 --max-insertion-length 5 --max-deletion-length 5 -g 3. All non-uniquely aligned reads were then removed. Owing to empirically determined mutation enrichment from non-template addition, 2 nt was trimmed from the 5′ end of each read. Mismatches located within 3 nt of an indel were also discarded for future analysis.

Bioinformatic analyses. Yeast genomic sequence conservation scores were obtained from *S. cerevisiae* genome database (https://www.yeastgenome.org). Positional coordinates of mRNAs including transcription start sites and polyadenylation sites are based on sacCer3 and reported measurement[40]. Genome-wide datasets were visualized by IGV software[41]. Structural models of the small and large ribosomal subunits including initiation factors were generated with the Chimera software[30,42]. Analyses of Gene Ontology term enrichment were carried out with GOrilla software using a single ranked list of genes[43].

RNA structure prediction was carried out with sequences 0–99 nt 3′ of the first nucleotide of an alternative start codon using the RNAfold web server[44]. Constraint settings were derived from DMS-MaPseq data as follows: a nucleotide was set as 'unpaired' if the DMS-MaPseq counts of a given nucleotide exceeded the value of 0.49 relative to the third highest count number in the range of 100 nt downstream of an ATIS.

Statistical significance of enrichment or depletion in certain regions (for example, Figs. 2c, 3d, e, Extended Data Fig. 10a) was determined by comparing weighted data vectors of the observed variable to the background value. To this end, we calculated *t*-values with the wtd.t.test function in R. The algorithm is based on the mean and 1/(standard errors)^2 as an estimate of the means accuracy. The given *P* values correspond to a two-tailed *t*-test.

Further bioinformatic analyses and multiple linear regressions were performed with R[45] with customized scripts using RStudio Software (https://www.rstudio.com/). Code is available upon request. Normalization of the datasets including ribosome protected fragments, monosome-protected fragments, Ded1p iCLIP–seq and DMS-MaPseq counts were performed relative to the total number of counts of the entire mRNA[42].

To compute Ded1p binding density, DMS-MapSeq ratios, or sequence conservation values in the vicinity of ATIS, it was important to normalize for inherent positional trends within the exact region in the respective iCLIP, DMS-MapSeq and sequence conservation datasets. For example, values for DMS-MapSeq ratios (counts *ded1-95*/ counts wild type), and iCLIP reads show an upward trend with increasing distance from the 5′ cap in 5′ UTRs. To normalize for inherent positional trends, we calculated a background distribution for the vicinity of each ATIS. We randomly choose a position in the respective section of a given mRNA, and determined the signal distribution in the vicinity of this position (for example, position −5 relative to the 5′ nt of an ATIS). This process was repeated four times. The background value reflects the average of these five calculated values. Reported enrichment values represent the ratio of the measured signal over the background value at each indicated position. Values are given in all plots as $\log_2$(measured signal/background signal). Statistical significance of enrichment or depletion was determined by calculating the *t*-value of the observed variable on the basis of the mean and s.d. of the background value.

Metagene profiles were calculated by averaging normalized Ded1p iCLIP counts and DMS- MaPseq counts after binning transcript coordinates from 5′ UTRs, ORFs and 3′ UTRs in bins reflecting 2% of each section of mRNA. Ded1p binding sites and the midpoint of RNA secondary structures were determined by Piranha peak calling software (http://smithlabresearch.org).

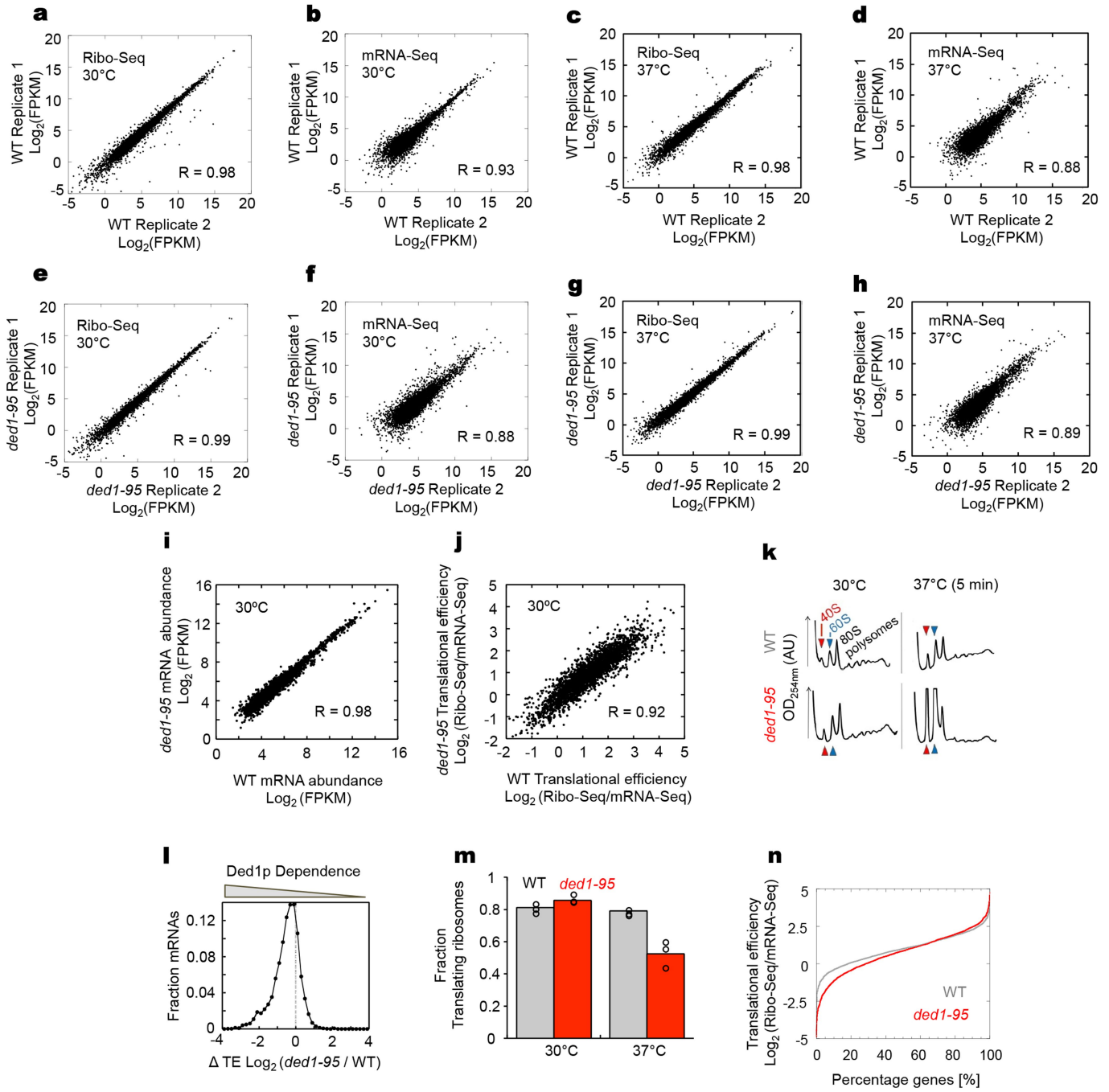
Calling parameters were optimized on the basis of visual inspection. To call peak sites of RNA secondary structures, a genome-wide dataset of $\log_2$(counts of DMS-MaPseq wild type/*ded1-95*) was used as input file.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) repository with the accession code GSE93959. All other data are available from the corresponding author upon reasonable request.

31. Hu, W., Sweet, T. J., Chamnongpol, S., Baker, K. E. & Coller, J. Co-translational mRNA decay in *Saccharomyces cerevisiae*. *Nature* **461**, 225–229 (2009).
32. Smith, J. E. et al. Translation of small open reading frames within unannotated RNA transcripts in *Saccharomyces cerevisiae*. *Cell Reports* **7**, 1858–1866 (2014).
33. Goecks, J., Nekrutenko, A. & Taylor, J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11**, R86 (2010).
34. Andreev, D. E. et al. Translation of 5′ leaders is pervasive in genes resistant to eIF2 repression. *eLife* **4**, e03971 (2015).
35. Matz, M. et al. Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res.* **27**, 1558–1560 (1999).
36. Nagalakshmi, U. et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349 (2008).
37. König, J. et al. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **17**, 909–915 (2010).
38. Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H. & Bartel, D. P. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508**, 66–71 (2014).
39. Licatalosi, D. D. et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).
40. Park, D., Morris, A. R., Battenhouse, A. & Iyer, V. R. Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. *Nucleic Acids Res.* **42**, 3736–3749 (2014).
41. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
42. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
43. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
44. Gruber, A. R., Bernhart, S. H. & Lorenz, R. The ViennaRNA web services. *Methods Mol. Biol.* **1269**, 307–326 (2015).
45. The R Core Team. *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria., http://www.R-project.org (2013).
46. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
47. Miyasaka, H. The positive relationship between codon usage bias and translation initiation AUG context in *Saccharomyces cerevisiae*. *Yeast* **15**, 633–637 (1999).
48. Nielsen, K. H. et al. Synergistic activation of eIF4A by eIF4B and eIF4G. *Nucleic Acids Res.* **39**, 2678–2689 (2011).
49. Floor, S. N., Condon, K. J., Sharma, D., Jankowsky, E. & Doudna, J. A. Autoinhibitory interdomain interactions and subfamily-specific extensions redefine the catalytic core of the human DEAD-box protein DDX3. *J. Biol. Chem.* **291**, 2412–2421 (2016).
50. Siepel, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).

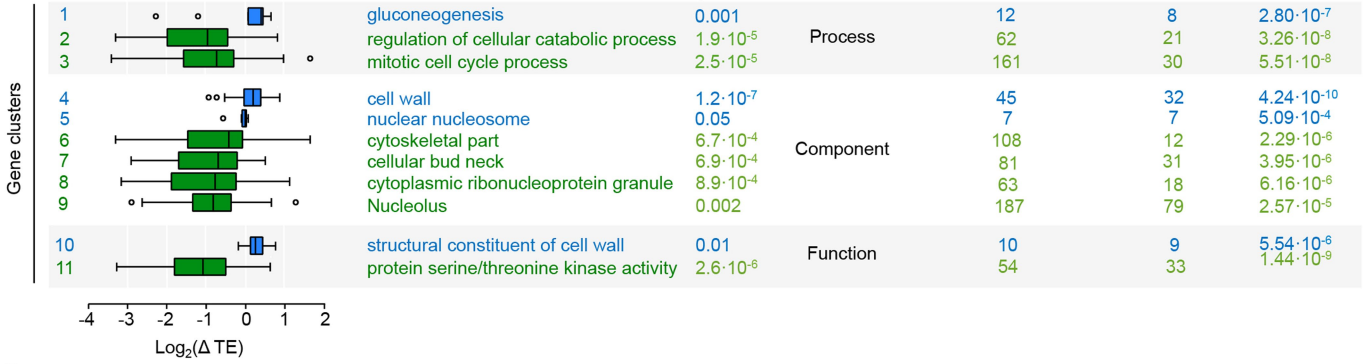
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 1 | mRNA expression, and translation profiles in wild-type *DED1* and *ded1-95*. **a**, Correlation of ribosome footprint counts between two biological replicates in wild-type *DED1* at 30 °C ($n = 5,523$). **b**, Correlation of mRNA expression levels between two biological replicates in wild-type *DED1* at 30 °C ($n = 5,372$). **c**, Correlation of ribosome footprint counts between two biological replicates in wild-type *DED1*, 5 min after temperature shift to 37 °C ($n = 5,523$). **d**, Correlation of mRNA expression levels between two biological replicates in wild-type *DED1*, 5 min after temperature shift to 37 °C ($n = 5,372$). **e**, Correlation of ribosome footprint counts between two biological replicates in *ded1-95* at 30 °C ($n = 5,523$). **f**, Correlation of mRNA expression levels between two biological replicates in *ded1-95* at 30 °C ($n = 5,372$). **g**, Correlation of ribosome footprint counts between two biological replicates in *ded1-95*, 5 min after temperature shift to 37 °C ($n = 5,523$). **h**, Correlation of mRNA expression levels between two biological replicates in *ded1-95*, 5 min after temperature shift to 37 °C ($n = 5,372$). **i**, Correlation of mRNA expression levels between wild-type *DED1* and *ded1-95* at 30 °C ($n = 2,976$). Each data point represents the average of at least two replicates. **j**, Correlation of translational efficiencies between wild-type *DED1* and *ded1-95* at 30 °C ($n = 2,976$). Each data point represents the average of at least two replicates. **k**, Representative polysome profiles of wild-type *DED1* and *ded1-95* strains at 30 °C and 5 min after shift to 37 °C. Similar results were obtained in three independent experiments. **l**, Changes in translational efficiencies (ΔTE) for mRNAs in *ded1-95*, compared to wild-type *DED1*, 5 min after temperature shift (mean of two biological replicates). The dotted line indicates no change. **m**, Fraction of 18S rRNA in polysome fractions, compared to the entire sample, at 30 °C and 5 min after temperature shift to 37 °C. Each bar represents the average of three independent experiments. Empty circles represent each replicate. **n**, Cumulative distribution of translational efficiencies of wild-type *DED1* and *ded1-95*, 5 min after temperature shift to 37 °C ($n = 2,976$). Each data point represents the average of at least two replicates.

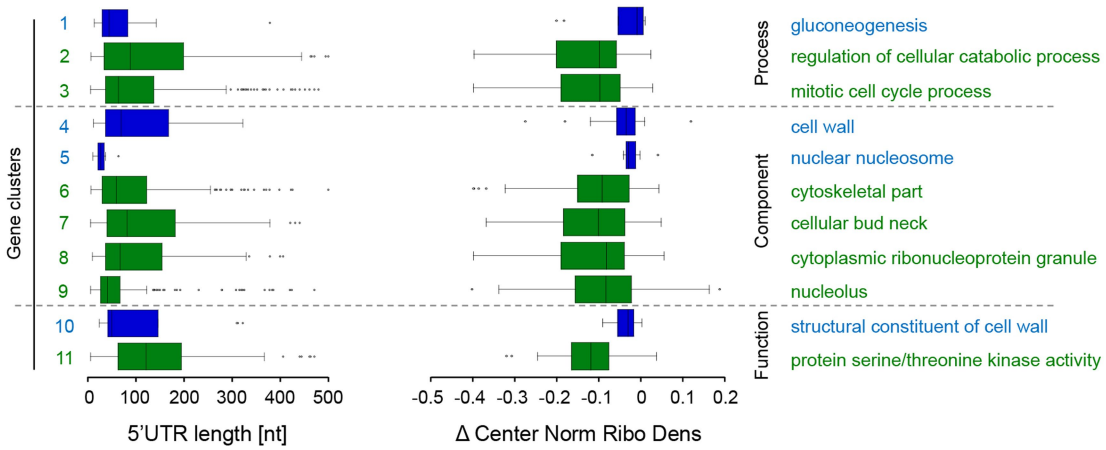
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

a



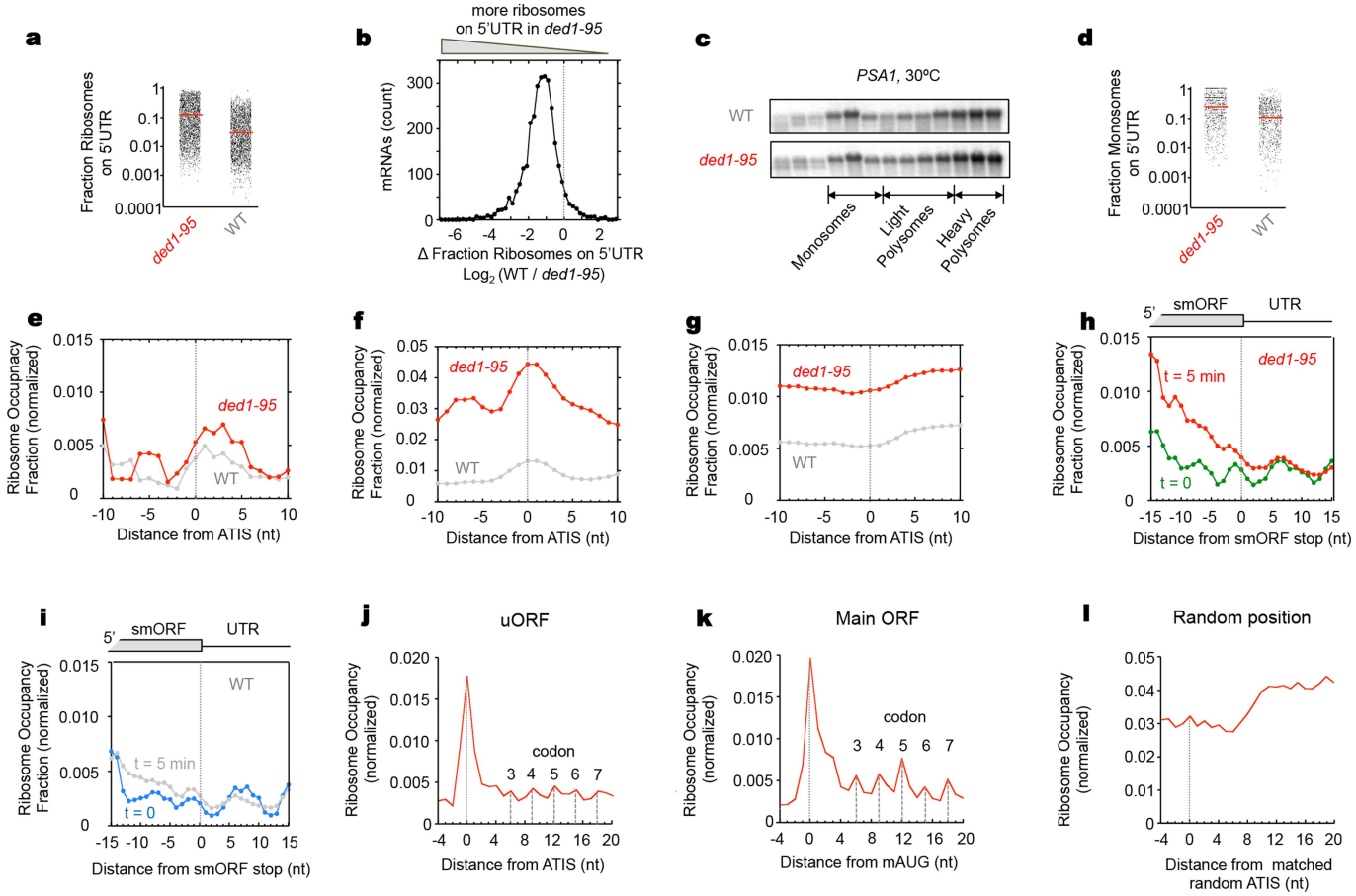
	GO term	FDR q-value	GO category	Number of genes with GO term	Number of genes enriched	P-value
1	gluconeogenesis	0.001	Process	12	8	$2.80 \cdot 10^{-7}$
2	regulation of cellular catabolic process	$1.9 \cdot 10^{-5}$		62	21	$3.26 \cdot 10^{-8}$
3	mitotic cell cycle process	$2.5 \cdot 10^{-5}$		161	30	$5.51 \cdot 10^{-8}$
4	cell wall	$1.2 \cdot 10^{-7}$	Component	45	32	$4.24 \cdot 10^{-10}$
5	nuclear nucleosome	0.05		7	7	$5.09 \cdot 10^{-4}$
6	cytoskeletal part	$6.7 \cdot 10^{-4}$		108	12	$2.29 \cdot 10^{-6}$
7	cellular bud neck	$6.9 \cdot 10^{-4}$		81	31	$3.95 \cdot 10^{-6}$
8	cytoplasmic ribonucleoprotein granule	$8.9 \cdot 10^{-4}$		63	18	$6.16 \cdot 10^{-6}$
9	Nucleolus	0.002		187	79	$2.57 \cdot 10^{-5}$
10	structural constituent of cell wall	0.01	Function	10	9	$5.54 \cdot 10^{-6}$
11	protein serine/threonine kinase activity	$2.6 \cdot 10^{-6}$		54	33	$1.44 \cdot 10^{-9}$

b



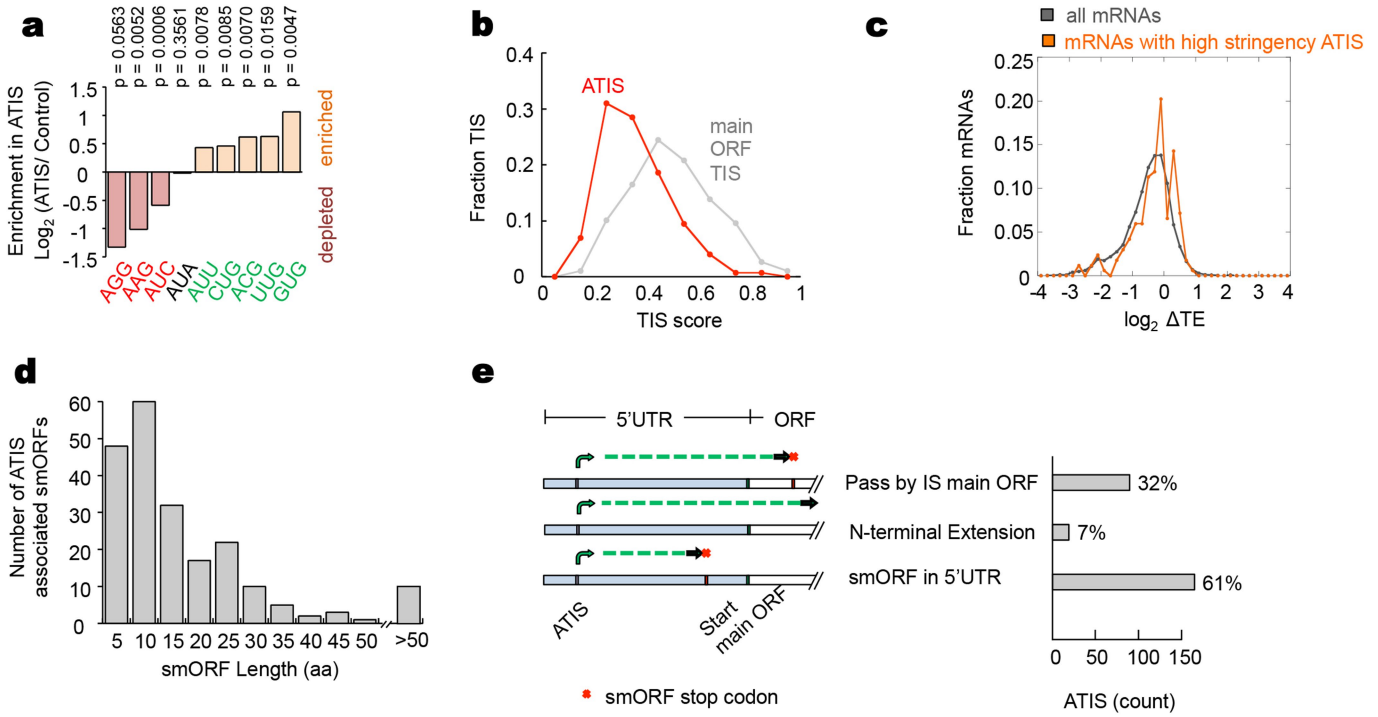
Extended Data Fig. 2 | A subset of mRNAs is largely insensitive to Ded1p. a, mRNA groups defined by Gene Ontology term with translation that is strongly affected (green) or largely unaffected by Ded1p (blue). Box plots (group median) of change in translational efficiencies. Box boundaries, upper and lower quartiles; error bars, 1.5 × interquartile range. The black box plot marks changes in translational efficiencies for all mRNAs (Fig. 1b). mRNAs for each Gene Ontology term were extracted from the *Saccharomyces* Genome Database (https://www.yeastgenome.org). The false discovery rate *q* value indicates the enrichment *P* value according to a hypergeometric model after correction for multiple testing using the Benjamini and Hochberg method[46]. **b**, Box plots (as in **a**) of 5′ UTR lengths and median of the shift in the normalized centre of ribosome density (Fig. 1b) for Gene Ontology term defined mRNA groups, colour-coded as in **a**.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



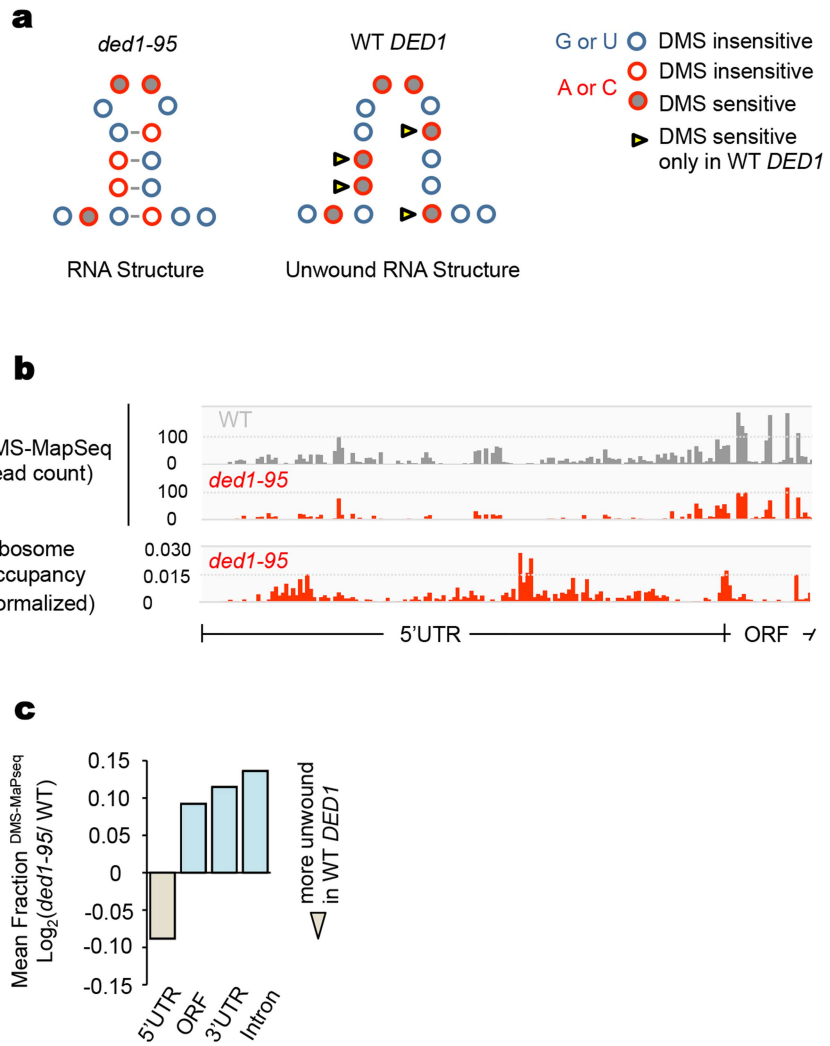
Extended Data Fig. 3 | Activation of ATIS in *ded1-95* upon temperature shift. a, Fraction of ribosome footprints on 5′ UTRs in wild-type *DED1* and *ded1-95*, (5 min, 37 °C, $n = 3{,}273$). The red line indicates the mean. Statistical significance for the difference between *ded1-95* and wild-type *DED1*: $P = 1.2 \times 10^{-119}$ (two tailed *t*-test). A similar result was obtained in an independent replicate ($P = 5.4 \times 10^{-47}$). **b**, Changes in the fraction of ribosomes on 5′ UTRs for all mRNAs ($n = 2{,}660$) in wild-type *DED1* compared to *ded1-95*, 5 min after temperature shift to 37 °C. The values on the *x*-axis represent the ratio ($\log_2$) of the fraction of ribosomes on each 5′ UTR in the wild type, divided by the fraction of ribosomes on the same 5′ UTR in *ded1-95*. Each value represents the mean of two independent biological replicates. **c**, Representative northern blots of PSA1 after sucrose gradient centrifugation for wild-type *DED1* and *ded1-95*, at 30 °C. A similar result was obtained in an independent biological replicate. **d**, Fraction of ribosome footprints on 5′ UTRs in wild-type *DED1* and *ded1-95*, (5 min, 37 °C), measured only in 80S monosomes ($n = 973$, reads from two independent experiments combined). Statistical significance for the difference between *ded1-95* and wild-type *DED1*: $P = 1.2 \times 10^{-50}$ (two tailed *t*-test). **e**, Mean ribosome occupancy within 10 nt 3′ and 5′ of the high-confidence ATIS in 5′ UTRs (moving average ± 1 nt, 5 min, 37 °C), measured without cycloheximide. **f**, Mean ribosome occupancy within 10 nt 3′ and 5′ of the high-confidence ATIS in 5′ UTRs ($n = 274$) in 80S monosomes (moving average ± 1 nt, 5 min, 37 °C). **g**, Mean ribosome occupancy within 10 nt 3′ and 5′ of all near-cognate initiation codons ($n = 61{,}614$; excluding medium-confidence ATIS) in 5′ UTRs (moving average ± 1 nt, 5 min, 37 °C). **h**, Ribosome occupancy of 3′ ends in small upstream open reading frames (smORFs) initiating at high-confidence ATIS in *ded1-95* before ($t = 0$) and after ($t = 5$ min) temperature shift. smORFs were included in this analysis if its length exceeds three codons and if the smORF terminates at least 11 nt upstream of the main AUG codon ($n = 76$). **i**, Ribosome occupancy of 3′ ends of smORFs defined in **e** in wild-type *DED1* before ($t = 0$) and after ($t = 5$ min) temperature shift. **j**, Ribosome occupancy 4 nt 5′ and 20 nt 3′ of high-confidence ATIS on 5′ UTRs ($n = 274$) for *ded1-95*, 5 min after temperature shift. The dashed lines indicate the first nucleotide of the marked in-frame codons. **k**, Ribosome occupancy 4 nt 5′ and 20 nt 3′ of the main AUG (mAUG) of mRNAs containing high-confidence ATIS on 5′ UTRs for *ded1-95*, 5 min after temperature shift. For mRNAs with multiple high-confidence ATIS in their 5′ UTR, the main AUG was counted only once. The dashed lines indicate the first nucleotide of the marked in-frame codons. **l**, Ribosome occupancy 4 nt 5′ and 20 nt 3′ of high-confidence ATIS-matched random positions (averaged from five randomizations) on 5′ UTRs ($n = 274$) for *ded1-95*, 5 min after temperature shift. The dashed line indicates the first nucleotide.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



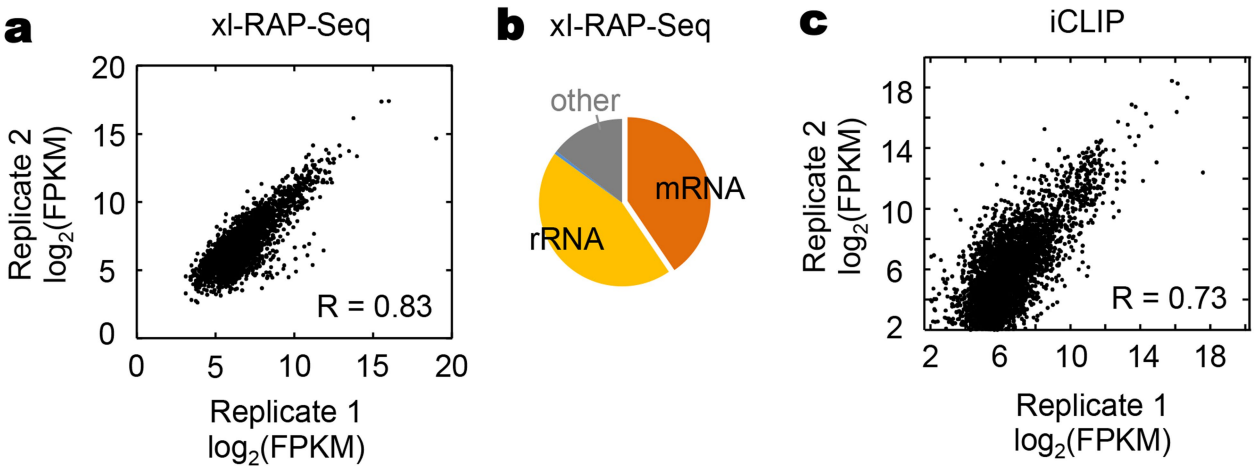
Extended Data Fig. 4 | Characteristics of small open reading frames associated with activated ATISs. a, Enrichment or depletion of each near-cognate codon in ATISs over the background distribution of the codon. *P* values determined using a two-tailed *t*-test. **b**, Mean translation initiation site score (positions −6 to +6, excluding +1 to +3), calculated according to previously published methods[47], for high-stringency ATIS ($n = 274$, red), and TIS of main ORFs ($n = 4{,}972$, grey). A TIS score exceeding 0.01 is considered a potential translational initiation site[14]. **c**, Changes in translational efficiencies (ΔTE) for mRNAs in *ded1-95*, compared to wild-type DED1, 5 min after temperature shift for all mRNAs (Fig. 1b) and ATIS-containing mRNAs. **d**, Length of the small open reading frames (smORFs) associated with *ded1-95*-activated ATIS. smORFs encoding N-terminal extensions were excluded from the analysis. **e**, Type of smORFs associated with *ded1-95*-activated ATIS. The bar graphs show the fraction of smORFs that falls into each category. The distribution of changes in translation efficiency (ΔTE) for RNAs with each type of smORF did not differ significantly.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



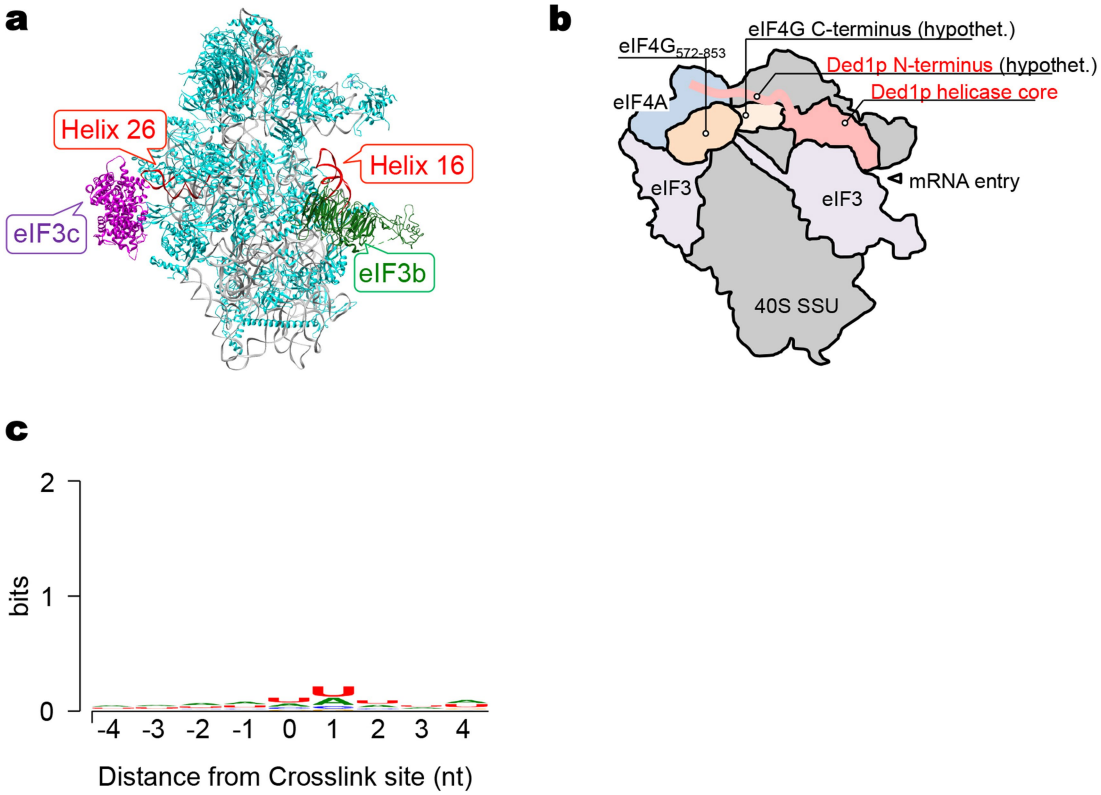
Extended Data Fig. 5 | mRNA structure unwinding by Ded1p in cells using DMS MaPSeq. a, Schematic for DMS-MaPseq approach to monitor RNA structure unwinding by Ded1p. All DMS-MapSeq experiments were performed 5 min after temperature shift. **b**, Representative DMS MaPSeq tracks in the *PSA1* 5′ UTR 5 min after temperature shift for wild-type *DED1* (grey) and *ded1-95* (red). Bars show normalized reverse transcription stops. A similar result was obtained in an independent replicate. The average Pearson's correlation coefficient of DMS-MaPseq counts per 5′ UTR between two replicates (5 min after temperature shift to 37 °C) were $R = 0.57$ ($n = 864$) for the wild type and $R = 0.63$ ($n = 692$) for *ded1-95*. The ribosome occupancy track for *ded1-95* is shown for reference. **c**, Unwinding of mRNA structure by Ded1p for different mRNA regions. Similar results were obtained in two independent experiments.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



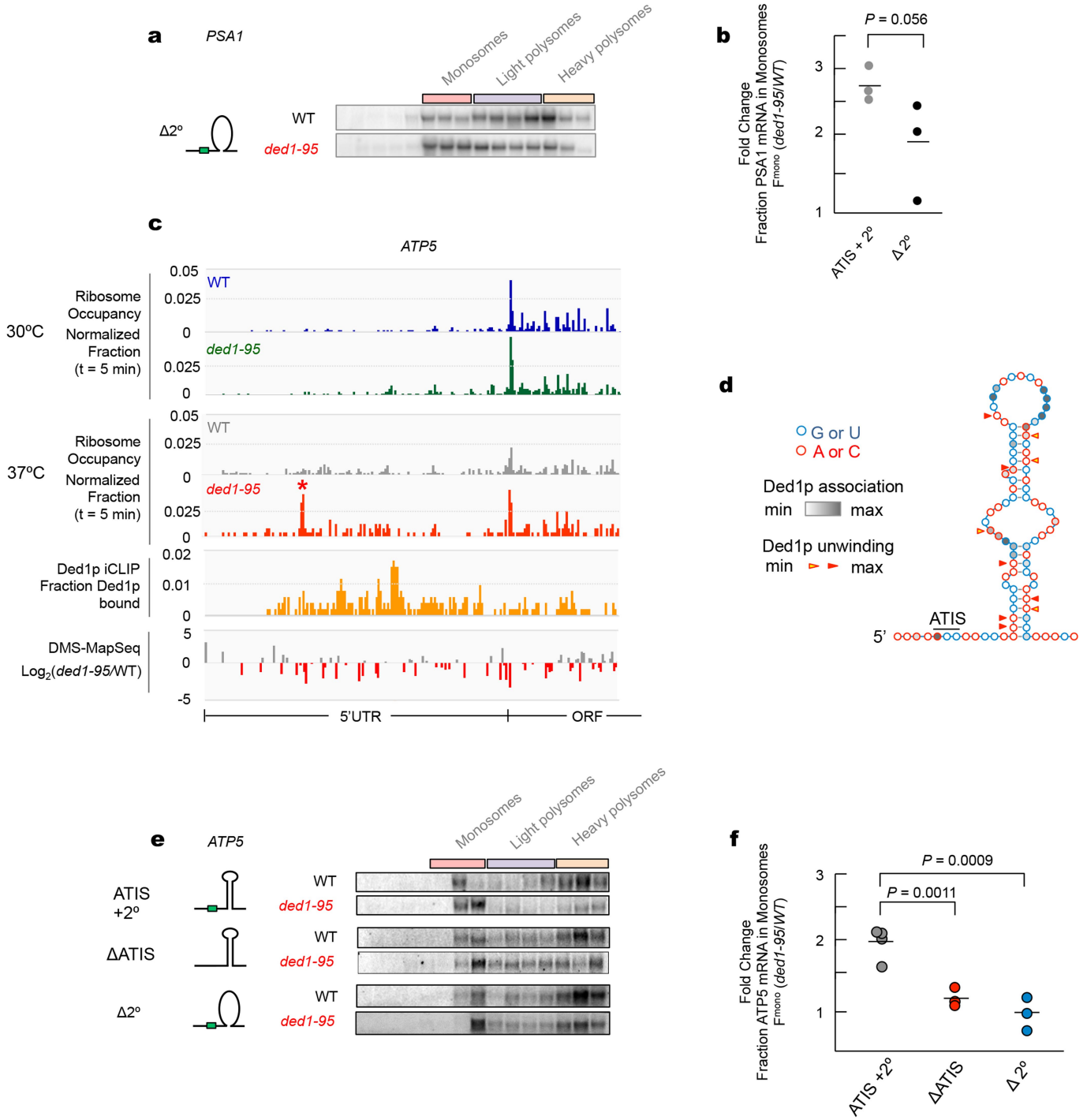
Extended Data Fig. 6 | XL-RAP–seq and iCLIP. **a**, Correlation of sequence reads (fragments per kilobase of exon per million fragments mapped, FPKM) per mRNA for two independent biological XL-RAP–seq replicates ($n = 2{,}992$) **b**, Fraction of mRNA (40%) and rRNA (44%) cross-linked to wild-type Ded1p as a fraction of all sequencing reads (mean of two independent experiments). $n = 4{,}280$ mRNAs exceed a minimal read count of FPKM ≥ 10. **c**, Correlation of the number of reverse transcription stops (FPKM) per mRNA for the two independent iCLIP approaches. Replicate 1, FLAG-tagged Ded1p; replicate 2, HTBH-tagged Ded1p; $n = 4{,}007$.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



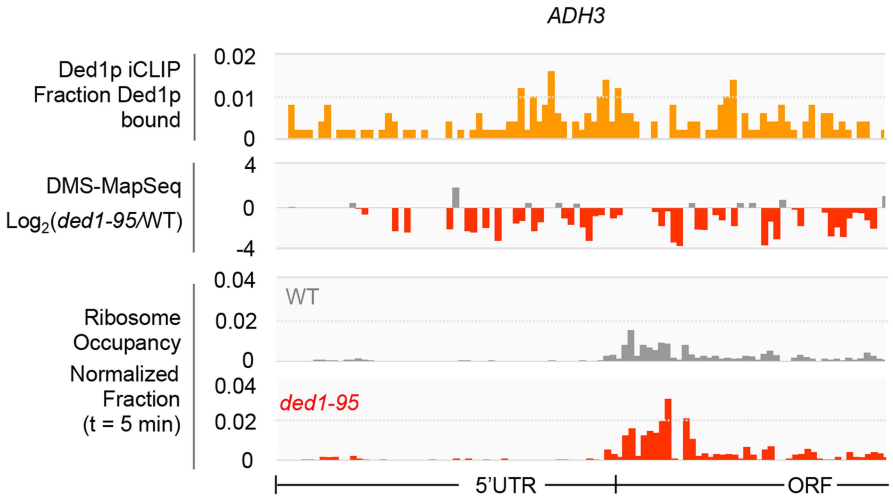
Extended Data Fig. 7 | Ded1p binding sites on 18S RNA and mRNAs. **a**, Ded1p binding sites on helix 720 (exit) and helix 16 (entry) (red) are in close proximity to the binding sites of eIF3c (purple) and eIF3b (green) on the 40S ribosomal subunit[42] (rRNA, grey; ribosomal proteins, cyan). **b**, Localization of Ded1p (apricot) on helix 16 of the PIC. Schematic model of the yeast PIC with eIF3 (http://www.bangroup.ethz.ch/research/eukaryotic_translation_initiation.html and references therein); the positioning of eIF4G$^{572-853}$ and eIF4A is derived from previously published work[48]. The position of the eIF4G C terminus is hypothetical. The helicase core of Ded1p was modelled in analogy to the DDX3 core structure[49], with the RNA binding site in contact with helix 16 at the main iCLIP cross-link sites. The position of the low-complexity N terminus of Ded1p is hypothetical. **c**, Sequence logo of Ded1p binding sites on mRNAs. Sets of 10^4 binding sites were randomly sampled from all Ded1p cross-linking sites and used as input to create a sequence logo (http://weblogo.berkeley.edu). All subsets yielded essentially the same sequence logo as shown here. Position zero denotes the reverse transcription stop.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



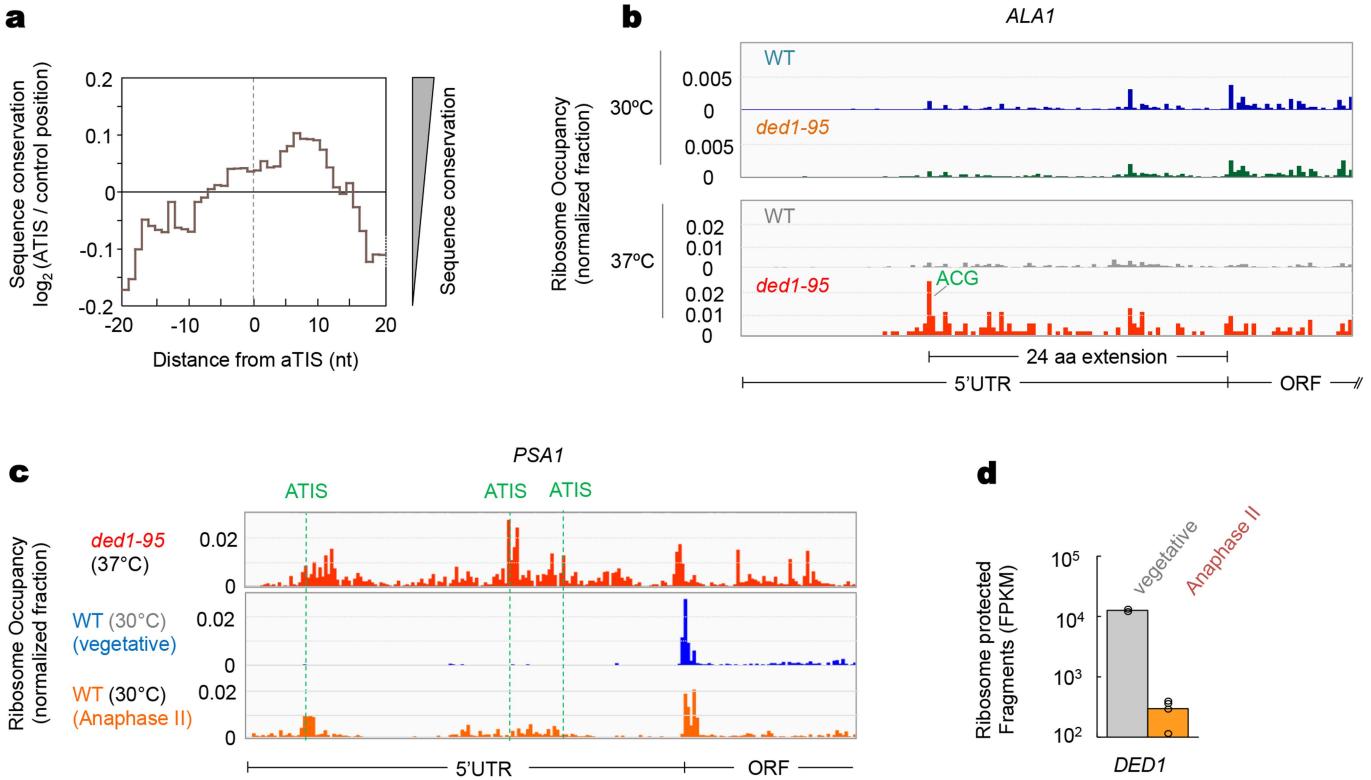
Extended Data Fig. 8 | Representative northern blots of PSA1 (Δ2°) and ATP5 (ΔATIS, Δ2°) for wild-type *DED1* and *ded1-95*. **a**, Representative RNA blots (5 min, 37 °C) for the *PSA1* mRNA with altered secondary structure, 3′ of the ATIS (Δ2°). Similar results were obtained in three independent biological replicates. **b**, Quantification of RNA blots for accumulation of the PSA1 Δ2° mRNA in monosomes in *ded1-95*, compared to wild-type *DED1*. The line indicates the average. The *P* value for the difference in monosome accumulation was determined using a one-tailed *t*-test. **c**, Representative ribosome profiling tracks for the 5′ UTR of *ATP5* in wild-type *DED1* and *ded1-95* (at 30 °C and after temperature shift). The near-cognate initiation codon is highlighted by a star. For comparison, Ded1p cross-linking (yellow track) and differential DMS-MaPseq tracks ($\log_2$(*ded1-95*/wild type), unwound mRNA regions are marked by red bars and have negative values) for the 5′ UTR of the *ATP5* mRNA are also shown. Similar results were obtained in two independent experiments. **d**, DMS-MapSeq-constrained secondary-structure model of a fragment of the *ATP5* mRNA 5′ UTR. The ATIS is marked by a line. Ded1p cross-linking (iCLIP) and unwinding (DMS-MapSeq) for each nucleotide are indicated. The ratio of normalized DMS-MapSeq counts of wild type/*ded1-95* in two categories: yellow triangles, 0.6 – 1.0 (moderately unwound) and red triangles, >1.0 (strongly unwound). **e**, Representative RNA blots (5 min, 37 °C) for wild-type *ATP5* mRNA and the same mRNA with mutations in the ATIS (ΔATIS) or with altered secondary structure 3′ of the ATIS (Δ2°) for wild-type *DED1* and *ded1-95*. Similar results were obtained in three independent experiments. **f**, Quantification of RNA blots for accumulation of the ATP5 ΔATIS and Δ2° mRNA in monosomes in *ded1-95*, compared to wild-type *DED1*. Lines indicate averages and *P* values were determined using a one-tailed *t*-test.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 9 | Ded1p binding and mRNA remodelling can occur without decreased translation efficiency if no near-cognate initiation codon is present. Ded1p iCLIP track, differential DMS-MaPseq track (5 min, 37 °C) and ribosome occupancy tracks (5 min, 37 °C) of wild-type *DED1* and *ded1-95* for *ADH3* mRNA, the translation of which is largely unaffected by Ded1p ($\Delta TE = -0.1$). 5′ UTR and ORF are marked. iCLIP and DMS-MaPseq tracks show Ded1p binding and remodelling of the 5′ UTR, ribosome profiling tracks indicate no significant accumulation of ribosomes in the 5′ UTR. Similar results were obtained in two independent experiments.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 10 | ATIS conservation across fungi and Ded1p-mediated activation of upstream ORFs starting from near-cognate initiation codons. a, Sequence conservation in fungi around high-confidence ATIS (moving average of ±1 nt). Positive values indicate higher sequence conservation than the average of five randomly chosen positions on the same 5′ UTR for each ATIS, negative values indicate less sequence conservation (conservation scores were obtained from the sacCer3 phastCons7way dataset, on the basis of sequence homology between the following species: *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, *S. bayanus*, *Naumoovozyma castellii* and *Lachancea kluyveri*)[50]. **b**, Ribosome occupancy tracks (30 °C and 5 min, 37 °C) of wild-type *DED1* and *ded1-95* for ALA1 mRNA. 5′ UTR and ORF are indicated. The ACG initiation codon (−25, marked) has been previously shown to function as an ATIS for the mitochondrial isoform of Ala1p[29]. Similar results were obtained in two independent biological replicates for each experiment. **c**, Ribosome occupancy tracks (5 min, 37 °C) of *ded1-95* and wild-type *DED1* (vegetative control and anaphase II) for *PSA1* mRNA. ATISs are marked by dashed lines. Similar results were obtained in two (vegetative control) and four (anaphase II) independent experiments. **d**, Ribosome-protected fragments mapping to *DED1* in vegetative cells and cells in anaphase II. Data are the mean of two (vegetative) and four (anaphase II) independent experiments, circles represent each replicate.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

https://doi.org/10.1038/s41586-018-0249-1

Cryo–EM structure of an essential *Plasmodium vivax* invasion complex

Jakub Gruszczyk[1,7], Rick K. Huang[2,7], Li-Jin Chan[1,3], Sébastien Menant[1], Chuan Hong[2], James M. Murphy[1,3], Yee-Foong Mok[4], Michael D. W. Griffin[4], Richard D. Pearson[5,6], Wilson Wong[1], Alan F. Cowman[1,3], Zhiheng Yu[2]* & Wai-Hong Tham[1,3]*

***Plasmodium vivax* is the most widely distributed malaria parasite that infects humans[1]. *P. vivax* invades reticulocytes exclusively, and successful entry depends on specific interactions between the *P. vivax* reticulocyte-binding protein 2b (*Pv*RBP2b) and transferrin receptor 1 (TfR1)[2]. TfR1-deficient erythroid cells are refractory to invasion by *P. vivax*, and anti-*Pv*RBP2b monoclonal antibodies inhibit reticulocyte binding and block *P. vivax* invasion in field isolates[2]. Here we report a high-resolution cryo-electron microscopy structure of a ternary complex of *Pv*RBP2b bound to human TfR1 and transferrin, at 3.7 Å resolution. Mutational analyses show that *Pv*RBP2b residues involved in complex formation are conserved; this suggests that antigens could be designed that act across *P. vivax* strains. Functional analyses of TfR1 highlight how *P. vivax* hijacks TfR1, an essential housekeeping protein, by binding to sites that govern host specificity, without affecting its cellular function of transporting iron. Crystal and solution structures of *Pv*RBP2b in complex with antibody fragments characterize the inhibitory epitopes. Our results establish a structural framework for understanding how *P. vivax* reticulocyte-binding protein engages its receptor and the molecular mechanism of inhibitory monoclonal antibodies, providing important information for the design of novel vaccine candidates.**

Invasion of red blood cells is an essential step in the life cycle of malaria parasites, and depends on interactions between parasite invasion ligands and their cognate red blood cell receptors[2–9]. *P. vivax* invades reticulocytes[10,11], and the *P. vivax* reticulocyte-binding protein (*Pv*RBP) family of adhesins has a crucial role in reticulocyte recognition[12–14]. A major pathway for *P. vivax* entry into reticulocytes is mediated by the specific interaction between a member of this protein family, *Pv*RBP2b, and the essential housekeeping protein TfR1[2].

We expressed and purified a recombinant fragment of *Pv*RBP2b encompassing residues 161–969 (*Pv*RBP2b(161–969)) in complex with TfR1 and transferrin (Tf), and reconstructed a three-dimensional electron density map using cryo-electron microscopy (cryo-EM) (Extended Data Fig. 1, Extended Data Table 1). The electron density corresponding to residues 161–167 and 634–969 could not be observed. The remaining residues, (*Pv*RBP2b(168–633)) adopt a highly elongated, predominantly α-helical fold that can be subdivided into two domains denoted as N- and C-terminal domains with an angle of ~130° between them (Fig. 1a). The N-terminal domain encompassing residues 168–460 and helices α1 to α7 can be superimposed onto the crystal structure of the corresponding fragment (PDB ID: 5W53)[2] with root mean square deviation (r.m.s.d.) of 0.84 Å over 279 aligned C_α atoms (Extended Data Fig. 2a). The structure of the C-terminal domain encompassing residues 461–633 was determined de novo and consists of a bundle of three long helices denoted as α8, α9 and α10 (Fig. 1a). The N-terminal domain is predominantly positively charged, in contrast to the negatively charged C-terminal domain (Extended Data Fig. 2b).

The cryo-EM structure of the *Pv*RBP2b–TfR1–Tf ternary complex was determined to an overall resolution of 3.7 Å (Fig. 1b, Extended Data Fig. 1). The size of the resolved structure is approximately 150×120×140 Å (Fig. 1b). The *Pv*RBP2b–TfR1–Tf complex consists

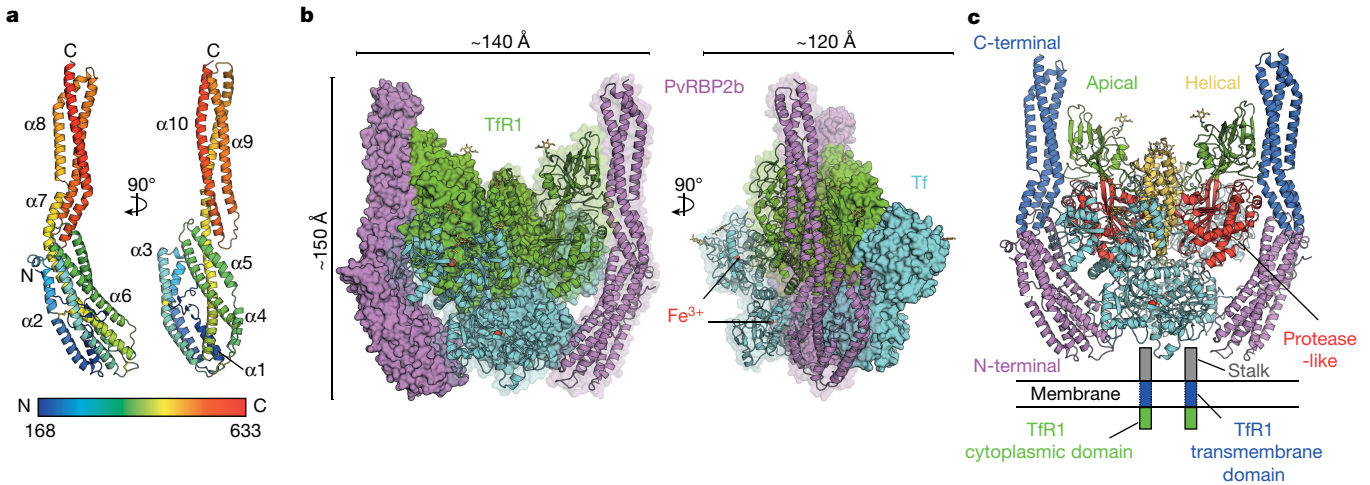


Fig. 1 | The cryo-EM structure of *Pv*RBP2b–TfR1–Tf ternary complex at 3.7 Å resolution. a, Structure of *Pv*RBP2b(168–633) in two orthogonal views with the secondary structure labelled. **b**, The two-ligand *Pv*RBP2b–TfR1–Tf ternary complex in two orthogonal views. TfR1, green; Tf, cyan; *Pv*RBP2b, violet; Ferric ions (Fe^{3+}), red spheres. **c**, The ternary complex relative to the reticulocyte membrane. *Pv*RBP2b: N-terminal domain, violet; C-terminal domain, blue. TfR1: apical domain, green; helical domain, yellow; protease-like domain, red. Tf, cyan; Fe^{3+}, red spheres.

[1]The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia. [2]CryoEM Shared Resource, Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. [3]Department of Medical Biology, The University of Melbourne, Melbourne, Victoria, Australia. [4]Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Melbourne, Victoria, Australia. [5]Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom. [6]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Oxford, United Kingdom. [7]These authors contributed equally: Jakub Gruszczyk, Rick K. Huang. *e-mail: yuz@janelia.hhmi.org; tham@wehi.edu.au

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

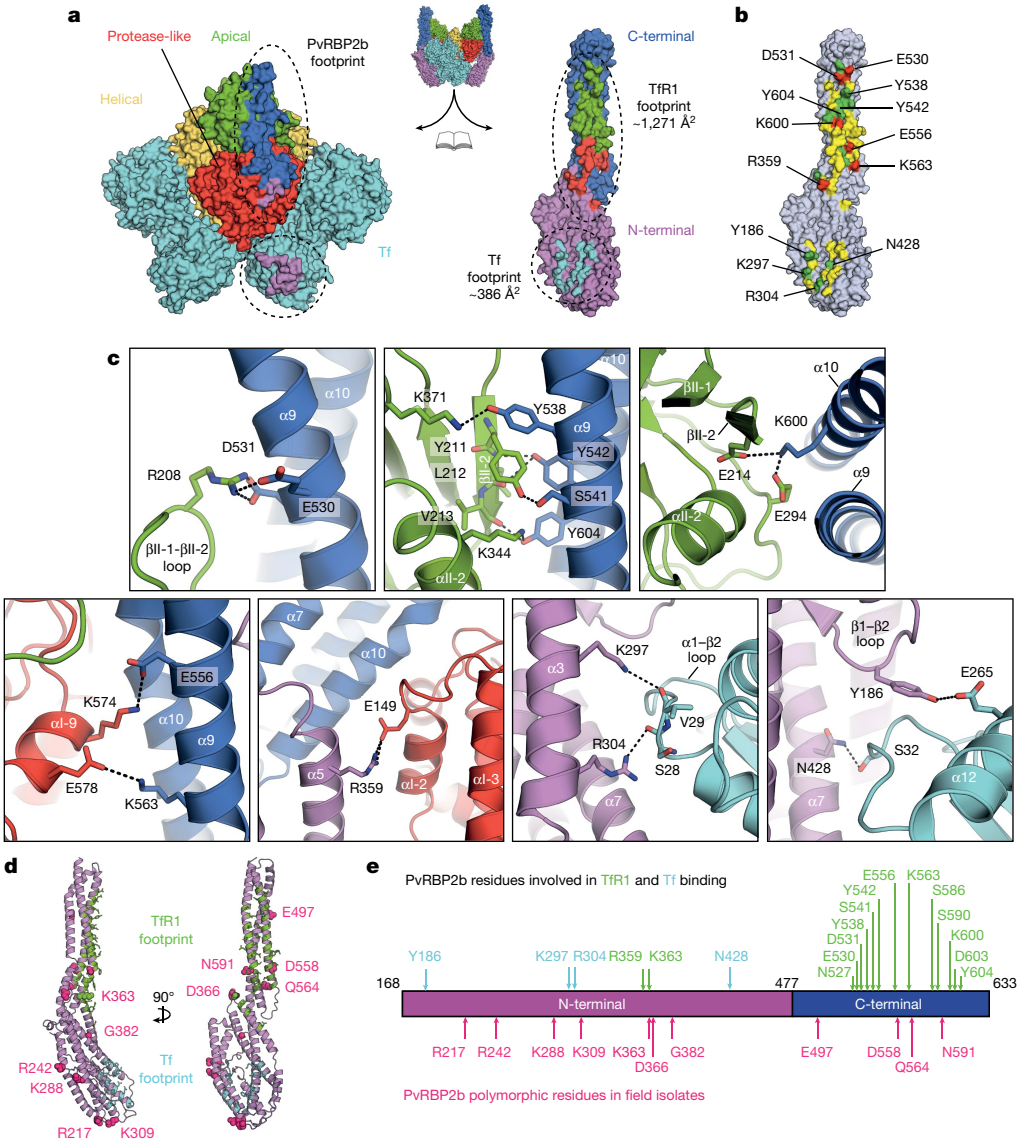


Fig. 2 | The *Pv*RBP2b–TfR1–Tf interface. a, Surface representation of the *Pv*RBP2b–TfR1–Tf complex. Left, TfR1: apical domain, green; helical domain, yellow; protease-like domain, red. Tf, cyan. TfR1 and Tf residues that interact with the N-terminal domain of *Pv*RBP2b are shown in violet; residues that interact with the C-terminal domain are shown in blue. Right, *Pv*RBP2b: N-terminal domain, violet; C-terminal domain, blue. *Pv*RBP2b residues that interact with Tf are shown in cyan; residues that interact with the apical domain of TfR1 are shown in green; residues that interact with the protease-like domain are shown in red. **b**, Residues of *Pv*RBP2b that bind TfR1 with hydrophobic interactions (yellow), hydrogen bonds (green) and salt bridges (red). **c**, Interaction sites between *Pv*RBP2b, TfR1 and Tf. Residues forming important contacts are shown as sticks and labelled. Proteins are coloured according to the domain organization as in **a**. **d**, Polymorphic residues labelled as pink spheres are mapped on the structure of *Pv*RBP2b, shown in two orthogonal views. Regions interacting with TfR1 and Tf are shown in green and cyan, respectively. **e**, Schematic of the *Pv*RBP2b sequence. Residues involved in binding and forming hydrogen bonds and salt bridges are indicated with arrows above the sequence and colour-coded as in **d**. Polymorphic residues are indicated with pink arrows below the sequence.

of the homodimeric TfR1 (residues 120–760) and its corresponding two molecules of iron-bound Tf (residues 1–679), with two molecules of *Pv*RBP2b bound on either side (Fig. 1c). We performed analytical ultracentrifugation experiments using a fixed amount of TfR1–Tf in the presence of increasing concentrations of *Pv*RBP2b (Extended Data Fig. 2c–e). These data are consistent with the binding of two molecules of *Pv*RBP2b to the TfR1–Tf complex with macroscopic dissociation constants in the mid-nanomolar to low-micromolar range. We determined another structure with only one *Pv*RBP2b protein bound to the TfR1–Tf binary complex, which allowed us to observe the conformational changes induced in the receptor upon ligand binding (Extended Data Fig. 3a–e). Additional classification further separated the two-ligand complex into two subclasses, which indicates an intrinsic flexibility of the N-terminal domain (Extended Data Fig. 3f, g). Overall, binding of *Pv*RBP2b results in small conformational changes that are confined to the apical domain of TfR1 and mostly involve rearrangement of the loops on the tip of the domain.

TfR1 is a single-pass type II transmembrane protein, with a cytoplasmic domain and a large extracellular domain[15]. The extracellular domain has a protease-like domain that resides proximal to the membrane, a helical domain that is responsible for dimerization, and an apical domain of unknown function (Fig. 1c). Co-crystal structures have been solved for human TfR1 in complex with Tf or HFE, or with the entry glycoprotein of a New World haemorrhagic arenavirus, Machupo virus (MACV GP1)[16–19]. The cryo-EM structure of the *Pv*RBP2b–TfR1–Tf complex shows that *Pv*RBP2b interacts with the TfR1–Tf binary complex through three principal sites: (i) the apical domain of TfR1, (ii) the protease-like domain of TfR1, and (iii) the N-terminal region of Tf (Fig. 2a–c, Extended Data Table 2). When bound to *Pv*RBP2b, TfR1 shows a more extensive buried surface

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

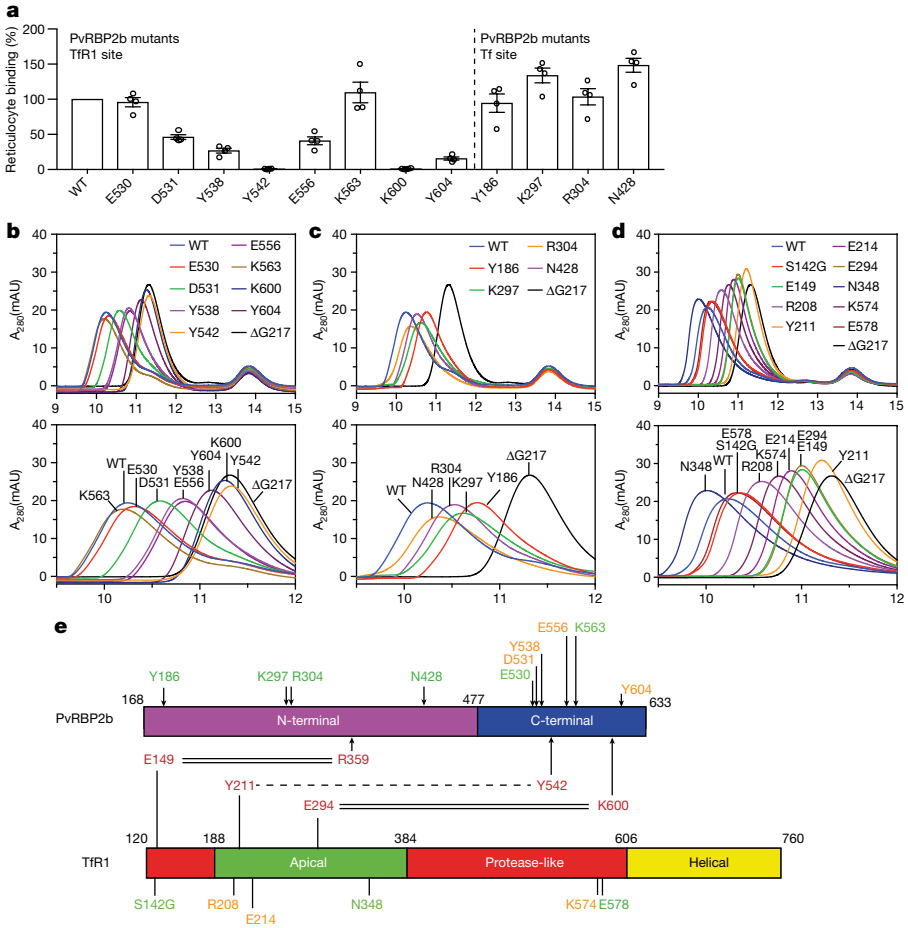


Fig. 3 | Structure–function analyses of *Pv*RBP2b and TfR1 residues involved in reticulocyte binding and complex formation. a, *Pv*RBP2b mutant proteins analysed using the reticulocyte-binding assay. Mean ± s.e.m, $n = 4$. Open circles, biological replicates. **b**, Elution profiles from analytical SEC, showing effects of mutations in *Pv*RBP2b residues that bind to TfR1. **c**, Analytical SEC showing effects of mutations in *Pv*RBP2b residues that bind to Tf. **d**, Analytical SEC showing effects of TfR1 mutations. **e**, Schematic diagrams summarizing the mutational analyses of *Pv*RBP2b and TfR1. Green, no defect; orange, moderate defect; red, complete defect. The key interactions between *Pv*RBP2b and TfR1 are highlighted as a dashed line for stacking interactions or double lines for salt bridges. All analytical SEC experiments in **b** and **d** were performed once.

area (approximately 1,271 Å^2) compared to Tf (approximately 386 Å^2) (Fig. 2a). We mapped naturally occurring field polymorphisms using data from the MalariaGEN *P. vivax* Genome Variation project onto the *Pv*RBP2b(168–633) structure[2,20] (Fig. 2d, e). Residues in *Pv*RBP2b that interact with TfR1 and Tf are mostly conserved, but reside within a region that is under balancing selection[2,20].

We performed site-directed mutagenesis to identify residues at the *Pv*RBP2b–TfR1 interface that are critical for complex formation (Extended Data Fig. 4a, b). We mutated the following residues on *Pv*RBP2b that interact with TfR1: R359, E530, D531, Y538, Y542, E556, K563, K600 and Y604, to alanine. These residues were selected as they are involved in either the formation of salt bridges or are part of the three-tyrosine motif (Fig. 2c). We also mutated four *Pv*RBP2b residues that interact with Tf, Y186, K297, R304 and N428, to alanine. Of the aforementioned mutants, *Pv*RBP2b(Y542A), *Pv*RBP2b(K600A) and *Pv*RBP2b(Y604A) reduced binding to reticulocytes by around 80% compared to wild-type *Pv*RBP2b (Fig. 3a).

We examined the ability of these *Pv*RBP2b mutant proteins to form a stable complex with TfR1–Tf using analytical size exclusion chromatography (SEC). As a control, we used TfR1 with a deletion in glycine 217 (TfR1(ΔG217)) which is known to abolish complex formation[2]. TfR1(ΔG217) is able to bind Tf, but did not form a ternary complex with *Pv*RBP2b (Fig. 3b–d). Of the *Pv*RBP2b mutants, only *Pv*RBP2b(Y542A), *Pv*RBP2b(K600A) and *Pv*RBP2b(Y604A) were defective in forming a complex with TfR1–Tf, as indicated by an overlap of elution profiles with TfR1ΔG217 (Fig. 3b, c). We examined whether stacking interactions mediated by the three-tyrosine motif at *Pv*RBP2b residues Y538, Y542 and Y604 were important for complex formation (Fig. 2c, top middle panel). Our results indicate that the hydrogen bonds formed by Y542 and the stacking interactions formed by aromatic side chains of Y538 and Y604 play an important role in the interaction of *Pv*RBP2b with TfR1 (Extended Data Fig. 4c). To examine the importance of the salt bridge at *Pv*RBP2b residue K600 (Fig. 2c, top right), we generated *Pv*RBP2b(K600E) and *Pv*RBP2b(K600M) mutants, in which the side chains are similar in length but lack the ability to form salt bridges. Both these mutants were defective in complex formation, supporting the importance of the salt bridge formed by the residue K600 (Extended Data Fig. 4c). These results indicate that *Pv*RBP2b residues Y542, K600 and Y604 are critical for reticulocyte-binding and complex formation.

We mutated the following residues on TfR1: E149, R208, Y211, E214, E294, K574 and E578, to alanine. We also modified S142 to glycine to mimic a common polymorphism in TfR1 that is prevalent in Asian populations, and N348 to alanine, as it is a critical determinant for MACV GP1 engagement with TfR1[17,21]. The overall structure of TfR1 is unperturbed by these single-site mutations, as shown by circular dichroism analyses and the observation that all TfR1 mutants retain their ability to bind Tf (Extended Data Fig. 4d–f). Of the aforementioned mutants, only TfR1(Y211A), TfR1(E294A) and TfR1(E149A) were unable to bind to *Pv*RBP2b, as indicated by an overlap of SEC elution profiles with TfR1(ΔG217) (Fig. 3d).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

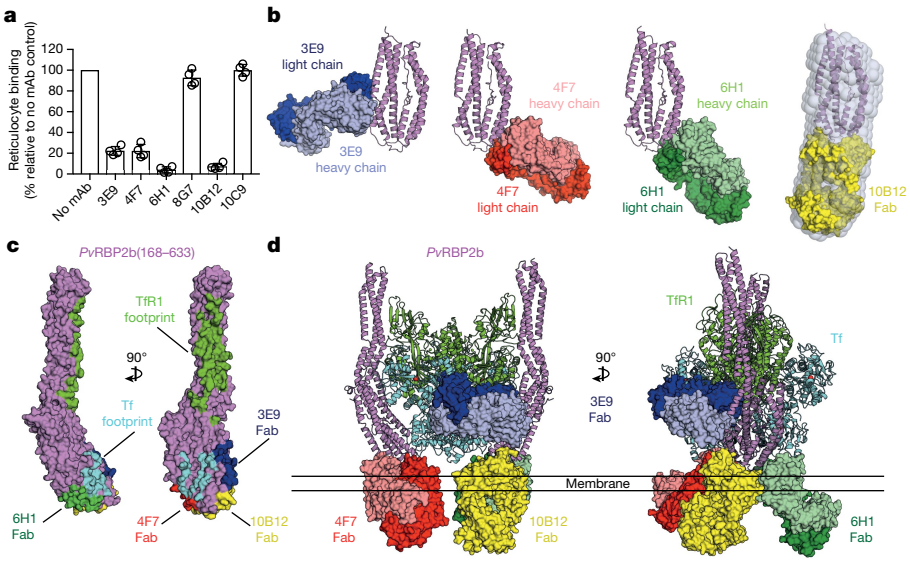


Fig. 4 | Structural modes of inhibition for anti-*Pv*RBP2b antibodies. **a**, *Pv*RBP2b binding to reticulocytes in the presence of anti-*Pv*RBP2b monoclonal antibodies 3E9, 4F7, 6H1, 8G7, 10B12, or anti-*Pf*Rh4 monoclonal antibody 10C9. Binding is normalized to *Pv*RBP2b binding in the absence of antibodies, which was arbitrarily assigned to be 100%. Mean ± s.e.m., $n = 4$; open circles, biological replicates. **b**, Crystal structures of *Pv*RBP2b (purple) bound to inhibitory antibody fragments of 3E9, 4F7 and 6H1, and SAXS-based model for the complex with 10B12 antibody fragment. **c**, Surface representation of the *Pv*RBP2b molecule with TfR1 (green), Tf (cyan) and the inhibitory-antibody interaction sites highlighted. **d**, Superposition of the structures of 3E9, 4F7, 6H1 and 10B12 antibody fragments on the cryo-EM structure of the *Pv*RBP2b–TfR1–Tf complex.

Our structure–function analyses highlight three key interactions that are important for complex assembly; those formed by *Pv*RBP2b(Y542) and TfR1(Y211), a salt bridge formed between *Pv*RBP2b(K600) and TfR1(E294), and a second salt bridge between TfR1(E149) and *Pv*RBP2b(R359) (Fig. 3e). Y211 is localized in the strand βII-2 in the apical domain of TfR1, and is a critical residue for entry of New World haemorrhagic arenaviruses[22]. Furthermore, TfR1(Y211) and TfR1(E149) are only present in human TfR1 but not in mouse TfR1, which explains why *P. vivax* does not invade rodent reticulocytes[21,23] (Extended Data Fig. 5a, b). We show that *Pv*RBP2b does not bind to TfR1–Tf binary complexes containing either mouse TfR1 or mouse Tf, indicating that *Pv*RBP2b displays specificity for human TfR1 and human Tf (Extended Data Fig. 5c). Whereas the majority of *Pv*RBP2b interactions involve TfR1, we hypothesize that interactions with Tf are important for complex stability, as only the *Pv*RBP2b–TfR1–Tf ternary complex could be isolated by size exclusion chromatography and visualized by cryo-EM techniques. These results show that *Pv*RBP2b binds to TfR1 residues that are not involved with Tf binding and identifies residues in TfR1 that determine *P. vivax* specificity for human reticulocytes.

The anti-*Pv*RBP2b monoclonal antibodies 3E9, 6H1 and 10B12 inhibit *Pv*RBP2b binding to reticulocytes and block the invasion of *P. vivax* into reticulocytes[2]. We identified an additional monoclonal antibody, 4F7, that inhibits *Pv*RBP2b binding to reticulocytes (Fig. 4a, Extended Data Fig. 6a). As expected, the non-inhibitory anti-*Pv*RBP2b antibody 8G7 and anti-PfRh4 antibody 10C9 did not affect *Pv*RBP2b binding[2] (Fig. 4a). We obtained crystal structures for *Pv*RBP2b–3E9, *Pv*RBP2b–4F7 and *Pv*RBP2b–6H1 complexes refined to 2.53, 2.66 and 3.34 Å, respectively, and derived a model for *Pv*RBP2b–10B12 from small angle X-ray scattering (SAXS) analysis (Fig. 4b–d, Extended Data Fig. 6b–f, Extended Data Tables 3, 4). 3E9 binds on the side of the N-terminal domain of *Pv*RBP2b to a region localized between helices α4, α5 and α6 (Fig. 4b, c). The binding sites for 6H1 and 4F7 antibody fragments overlap; they bind close to the tip of the N-terminal domain of *Pv*RBP2b, forming interactions with residues localized in helix α2 and α3 as well as to the loop connecting helices α3 and α4 (Fig. 4b, c). 10B12 binds at the tip of *Pv*RBP2b at a distinct site from 6H1 and 4F7 (Fig. 4b, c). When superimposed on the ternary complex structure (Fig. 4c, d), 3E9 sterically clashes with TfR1 and Tf. Superimposition of crystal structures of 4F7, 6H1 and 10B12 onto the ternary complex structure shows that these antibodies do not inhibit by blocking the sites of interaction with TfR1 or Tf, but do so through steric hindrance with the reticulocyte membrane (Fig. 4d). Whereas all these antibodies target the N-terminal domain of *Pv*RBP2b, future work is needed to determine whether antibodies against the C-terminal domain will be effective at blocking interactions of *Pv*RBP2b with TfR1.

Cryo-EM single-particle analysis of the *Pv*RBP2b-TfR1-Tf complex identified critical interfaces between *Pv*RBP2b and TfR1 that are required for complex formation. *P. vivax* harbours more global genetic diversity relative to *P. falciparum*; however, the *Pv*RBP2b residues that are critical for complex formation are conserved across more than 200 field isolates from Asia Pacific[20,24]. The crystal structures of the inhibitory antibody fragments with *Pv*RBP2b provide mechanistic insights into the modes of action that involve steric hindrance with either the receptor or the reticulocyte membrane. These high-resolution structures provide a solid structural framework for rational design of candidates vaccines to block *P. vivax* invasion into reticulocytes.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0249-1.

Received: 14 December 2017; Accepted: 1 May 2018;
Published online: 27 June 2018

1. Mueller, I. et al. Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite. *Lancet Infect. Dis.* **9**, 555–566 (2009).
2. Gruszczyk, J. et al. Transferrin receptor 1 is a reticulocyte-specific receptor for *Plasmodium vivax*. *Science* **359**, 48–55 (2018).
3. Tham, W.-H. et al. Complement receptor 1 is the host erythrocyte receptor for *Plasmodium falciparum* PfRh4 invasion ligand. *Proc. Natl Acad. Sci. USA* **107**, 17327–17332 (2010).
4. Horuk, R. et al. A receptor for the malarial parasite *Plasmodium vivax*: the erythrocyte chemokine receptor. *Science* **261**, 1182–1184 (1993).
5. Miller, L. H., Mason, S. J., Dvorak, J. A., McGinniss, M. H. & Rothman, I. K. Erythrocyte receptors for (*Plasmodium knowlesi*) malaria: Duffy blood group determinants. *Science* **189**, 561–563 (1975).
6. Adams, J. H. et al. The Duffy receptor family of *Plasmodium knowlesi* is located within the micronemes of invasive malaria merozoites. *Cell* **63**, 141–153 (1990).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

7. Cowman, A. F., Tonkin, C. J., Tham, W.-H. & Duraisingh, M. T. The molecular basis of erythrocyte invasion by malaria parasites. *Cell Host Microbe* **22**, 232–245 (2017).
8. Maier, A. G. et al. *Plasmodium falciparum* erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. *Nat. Med.* **9**, 87–92 (2003).
9. Crosnier, C. et al. Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum*. *Nature* **480**, 534–537 (2011).
10. WHO. World Malaria Report 2016. *WHO* http://www.who.int/malaria/publications/world-malaria-report-2016/report/en/. (2017)
11. Hegner, R. Relative frequency of ring-stage plasmodia in reticulocytes and mature erythrocytes in man and monkey. *Am. J. Epidemiol.* **27**, 690–718 (1938).
12. Carlton, J. M. et al. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* **455**, 757–763 (2008).
13. Galinski, M. R., Medina, C. C., Ingravallo, P. & Barnwell, J. W. A reticulocyte-binding protein complex of *Plasmodium vivax* merozoites. *Cell* **69**, 1213–1226 (1992).
14. Gruszczyk, J. et al. Structurally conserved erythrocyte-binding domain in *Plasmodium* provides a versatile scaffold for alternate receptor engagement. *Proc. Natl Acad. Sci. USA* **113**, E191–E200 (2016).
15. Lawrence, C. M. et al. Crystal structure of the ectodomain of human transferrin receptor. *Science* **286**, 779–782 (1999).
16. Lebron, J. A. et al. Crystal structure of the hemochromatosis protein HFE and characterization of its interaction with transferrin receptor. *Cell* **93**, 111–123 (1998).
17. Abraham, J., Corbett, K. D., Farzan, M., Choe, H. & Harrison, S. C. Structural basis for receptor recognition by New World hemorrhagic fever arenaviruses. *Nat. Struct. Mol. Biol.* **17**, 438–444 (2010).
18. Radoshitzky, S. R. et al. Transferrin receptor 1 is a cellular receptor for New World haemorrhagic fever arenaviruses. *Nature* **446**, 92–96 (2007).
19. Cheng, Y., Zak, O., Aisen, P., Harrison, S. C. & Walz, T. Structure of the human transferrin receptor-transferrin complex. *Cell* **116**, 565–576 (2004).
20. Pearson, R. D. et al. Genomic analysis of local variation and recent evolution in *Plasmodium vivax*. *Nat. Genet.* **48**, 959–964 (2016).
21. Demogines, A., Abraham, J., Choe, H., Farzan, M. & Sawyer, S. L. Dual host-virus arms races shape an essential housekeeping protein. *PLoS Biol.* **11**, e1001571 (2013).
22. Radoshitzky, S. R. et al. Receptor determinants of zoonotic transmission of New World hemorrhagic fever arenaviruses. *Proc. Natl Acad. Sci. USA* **105**, 2664–2669 (2008).
23. Mikolajczak, S. A. et al. *Plasmodium vivax* liver stage development and hypnozoite persistence in human liver-chimeric mice. *Cell Host Microbe* **17**, 526–535 (2015).
24. Neafsey, D. E. et al. The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. *Nat. Genet.* **44**, 1046–1050 (2012).

Acknowledgements We thank J. Newman from CSIRO Collaborative Crystallization Centre for assistance with setting up the crystallization screens, the Walter and Eliza Hall Institute's Monoclonal Antibody Facility for production of antibodies, and the MX and SAXS beamline staff at the Australian Synchrotron for their assistance during data collection. This work was supported in part by the Australian Research Council Future Fellowship to W.-H.T. and M.D.W.G., a Speedy Innovation Grant to W.-H.T. and a National Health and Medical Research Council fellowship (1105754) to J.M.M. W.-H.T. is a Howard Hughes Medical Institute–Wellcome Trust International Research Scholar (208693/Z/17/Z). R.D.P. is funded by Wellcome Trust (090770). The authors acknowledge support from the Drakensberg Trust, the Victorian State Government Operational Infrastructure Support and Australian Government NHMRC IRIISS.

Reviewer information *Nature* thanks S. Scheres and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions J.G., L.-J.C. and S.M. established experimental procedures. J.G. and W.W. optimized the samples for cryo-EM. R.K.H., C.H. and Z.Y. performed specimen preparation for cryo-EM, collected and processed the data and generated the maps. J.G. carried out model building, refinement and analysed the structure of *Pv*RBP2b–TfR1–Tf complex. J.G., L.-J.C. and S.M. expressed and purified all mutants and performed the structure–function analyses. L.-J.C. and J.G. purified the Fab fragments. J.G. crystallized, collected data and solved the structures of Fab complexes. J.M.M. and J.G. collected and analysed SAXS data. Y.-F.M. and M.D.W.G. collected and analysed AUC data. R.D.P. performed the sequence diversity analyses. W.-H.T., J.G. and A.F.C. designed the project. W.-H.T., J.G. and Z.Y. analysed the data, and W.-H.T. and J.G. principally wrote the manuscript which was finalized with input from all authors.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0249-1.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0249-1.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to Z.Y. or W.-H.T.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Expression and purification of recombinant *Pv*RBP2b and TfR1 mutants. Protein expression and purification for wild-type *Pv*RBP2b, wild-type TfR1 and Tf were performed as described previously[2]. Mouse Tf was purchased from Sigma–Aldrich. Restriction-free cloning was used to generate the alanine mutants of *Pv*RBP2b and TfR1, and all sequences were verified at the Melbourne Translational Genomics Center. Expression and purification of all mutant proteins were performed as for the wild-type proteins. We were unsuccessful in multiple purification attempts for PvRBP2b(R359A). The complex between *Pv*RBP2b, TfR1 and Tf was prepared by mixing equimolar amounts of each protein to a final concentration of 1.0 mg/ml in buffer containing 20 mM NaHEPES pH 7.5, 100 mM NaCl and 50 mM $NaHCO_3$. The sample was flash frozen in liquid nitrogen after the addition of 10% (v/v) glycerol, and stored at −80 °C. Circular dichroism experiments were performed as described previously[2] to compare the secondary structure of wild-type versus mutant proteins.

Cryo-EM sample preparation and data acquisition. Prior to grid preparation, an aliquot of protein was thawed on ice, immediately followed by glycerol removal using a 0.5 μl 100-kDa MWCO Amicon filtration unit (Millipore) at 4 °C and 2,000 *g* for 5 cycles. Sample (3.2 μl) in glycerol-free buffer (20 mM NaHEPES pH 7.5, 100 mM NaCl and 50 mM $NaHCO_3$) was applied to a glow-discharged 400-mech quantifoil 1.2/1.3 Au grid (Quantifoil, Großlöbichau Germany) and then rapidly plunge-frozen into a liquid ethane bath on a Vitrobot (FEI company, part of Thermo Fisher Scientific). Data were acquired at a sampling rate of 1.35 Å/pixel on a FEI Titan Krios at 300 kV equipped with spherical aberration corrector and a post-energy filter K2 direct electron detector. Each exposure was recorded with a total dose of ~80 electron/$Å^2$ at a frame rate of 0.3 s per frame.

Image processing and reconstruction. Motion and drift in each image stack was measured, corrected, and dose-weighted by unblur and sum_movie at 1.65 electron/$Å^2$ per frame[25]. Three thousand particles were manually picked in EMAN2[26] followed by 2D class averages calculation in Relion. The preliminary 2D class averages were used to generate templates for automated particle picking. CTF determination for each drift-corrected micrograph was calculated by CTFFIND4[27]. We used Relion1.4 and Relion2.0[28] for particle picking, extraction, 2D and 3D classifications, and high-resolution refinement to derive final reconstructions for one-ligand and two-ligand bound complexes. 3D classification was performed without imposing any symmetry. As a result, one-ligand and two-ligand subclasses were identified. One-ligand complex continued to be refined without symmetry being applied. The two-ligand class was refined independently in both C1 and C2 symmetry, and both refinements yielded nearly identical high-resolution reconstructions (3.74 Å versus 3.68 Å). The difference between C1 and C2-applied maps is negligibly minor, indicating that transient heterogeneity was not detected under the current resolution limit. Therefore, C2 symmetry was applied in local refinements and post-processing to generate the final reconstructed map of the two-ligand complex.

Model building and refinement. The program Chimera[29] was initially employed to dock the available crystal structures into the electron density maps obtained from cryo-EM. We used the following models: for TfR1, PDB ID 1CX8[15]; for Tf, PDB ID 3QYT[30]; and for the N-terminal domain of *Pv*RBP2b (residues 169–470), PDB ID 5W53[2]. For Tf, we observed some conformational differences between the model and electron density, and to facilitate the fitting we divided the molecule into three parts, consisting of subdomains N1, N2 and the C lobe. Model building and refinement was continued manually in Coot[31]. The C-terminal domain of *Pv*RBP2b (residues 471–633) was built de novo. The high resolution of this region allowed unambiguous assignment of the side chains in the region that interacts with TfR1. We did not observe the electron density corresponding to residues 161–167 and 634–969.

The model of the ternary complex was refined in Phenix[32] using real space refinement. For the two-ligand complex, non-crystallographic symmetry (NCS) was imposed on the corresponding chains of identical molecules. For one-ligand complex, no NCS restrains were used. No secondary structure restraints were initially applied to the model. The final assignment of secondary structures was performed using the DSSP program[33] integrated in Chimera[29]. The quality of the final models was validated using MolProbity[34] and EMRinger[35]. Electron density maps were visualized using Chimera[29].

Determination of field polymorphisms. We determined naturally occurring field polymorphisms using data from the May 2016 release of the MalariaGEN *P. vivax* Genome Variation project. We selected all variants that passed filters and had a non-reference allele frequency among all samples of >0.1. These variants can be accessed at https://www.malariagen.net/apps/pvgv/index.html?dataset=pvivax_stable&workspace=workspace_1&view=cc42817a-d8dd-11e7-a74a-22000a4d9adb&state=table_variants.

Antibody production. Mouse monoclonal antibodies against *Pv*RBP2b were prepared at the Walter and Eliza Hall Institute Monoclonal Antibody Facility as described previously[2]. Twenty-four hybridomas were selected, on the basis of displaying the strongest antibody response to *Pv*RBP2b(161–1454) as measured by ELISA. These hybridoma supernatants were screened for the ability to inhibit *Pv*RBP2b(161–1454) binding to reticulocytes. Selected hybridomas producing inhibitory antibodies were cloned by limiting dilution in multiwell plates, aiming for one cell or less per well. The sub-cloned cell supernatants were screened by ELISA against *Pv*RBP2b(161–1454). Two or more rounds of limiting dilution cloning were generally required before the hybridomas were deemed monoclonal. The antibodies were purified from monoclonal hybridoma supernatants with protein A sepharose. All monoclonal antibodies used for protein crystallization have been sequenced by GenScript.

Flow cytometry-based reticulocyte-binding assay. Enriched reticulocytes were resuspended in phosphate buffered saline (PBS) to a final volume of 1×10^7 cells/ml. Recombinant proteins were incubated at 0.02 mg/ml in 50 μl of the resuspended reticulocytes for 1 h at room temperature. All washes were performed in PBS supplemented with 1% (w/v) bovine serum albumin (BSA) and spun at 4,000*g* for 1 min. All antibody incubations were performed at room temperature for 1 h. Binding assays were washed once and incubated with the respective rabbit polyclonal antibodies (12.5 μg/ml). After washing, Alexa Fluor 647 chicken anti-rabbit secondary antibody (1:100; Life Technologies) was added. After a final wash, 100 μl thiazole orange (Retic-Count Reagent; BD Biosciences) was added and incubated for half an hour. Reticulocytes with bound proteins were resuspended in 200 μl PBS and analysed on the LSR II flow cytometer (BD Biosciences). 50,000 events were recorded and results were analysed using FlowJo software (Three Star). The background signal from a rabbit polyclonal antibody and Alexa Fluor 647-conjugated antibody control (without protein) was subtracted from all binding results. To enable comparison between biological repeats, the percentage binding of the protein in the presence of mAbs was divided by the percentage binding with no mAbs and multiplied by 100 to obtain the percentage binding relative to the no inhibitor control.

Analytical size exclusion chromatography. Liquid chromatography was performed using ÄKTA pure 25 M1 chromatographic system (GE Healthcare). Proteins were mixed in equimolar ratio 1 h before the analysis and the same amount of protein was used throughout the experiment. 100 μl was injected onto Superdex 200 Increase 10/300 GL column (GE Healthcare) equilibrated with buffer containing 20 mM NaHEPES pH 7.5, 100 mM NaCl, 50 mM $NaHCO_3$ at 0.75 ml/min of buffer flow. The absorbance of the eluent was monitored at 280 nm. Eluates were collected in 0.5 ml fractions and analysed using SDS–PAGE. All steps were performed at room temperature.

Analytical ultracentrifugation. Sedimentation velocity experiments were performed using an XL-I analytical ultracentrifuge (Beckman Coulter) equipped with UV/Vis scanning optics. Samples were prepared containing a constant concentration of 2.5 μM TfR1–Tf complex and *Pv*RBP2b(161–969) at concentrations of 0.25, 0.5, 1.0, 2.0, 4.0, 8.0 and 16.0 μM. Buffer reference (20 mM NaHEPES pH 7.5, 100 mM NaCl, 50 mM $NaHCO_3$) and sample solutions were loaded into 12-mm double-sector cells with quartz windows and the cells were mounted in an An-50Ti 8-hole rotor. Centrifugation was conducted at 35,000 r.p.m. (98,780 *g*) at 20 °C, and radial absorbance data were collected at 250 nm in continuous mode. Data were fitted to a continuous sedimentation coefficient distribution ($c(s)$) model and distributions were integrated between 9 *S* and 13 *S* using SEDFIT[36] to determine the integrated absorbance signal. Buffer density and viscosity were calculated using SEDNTERP[37].

Fab fragment purification and formation of Fab–*Pv*RBP2b complexes. Mouse monoclonal antibodies were digested using papain (Sigma–Aldrich) for 2 h at room temperature. For 6H1 and 10B12, L-cysteine (Sigma–Aldrich) was added to cleave $F(ab')_2$ fragments. An excess of iodoacetamide (Sigma–Aldrich) was added to halt the reaction and the samples were passed through a protein A column (GE Healthcare). The flow through was concentrated and applied on a S75 Superdex 16/600 column (GE Healthcare) equilibrated with a buffer containing 20 mM NaHEPES pH 7.5 and 150 mM NaCl. Protein samples were flash frozen in liquid nitrogen after addition of 10% (v/v) glycerol, and stored at −80 °C until further processing.

Recombinant *Pv*RBP2b fragment *Pv*RBP2b(169–470) was mixed with molar excess of individual Fab fragments and incubated overnight at 4 °C to allow complex formation and purified using SEC. Fractions containing the complexes were concentrated using Vivaspin 15 Turbo centrifugal concentrators (Sartorius) with a 5-kDa molecular weight cut-off in a buffer containing 20 mM NaHEPES pH 7.5 and 150 mM NaCl and used for crystallization.

Protein crystallization. Crystallization trials were performed at the CSIRO Collaborative Crystallization Center using two different protein concentrations, 7.5 and 15 mg/ml. Crystals of 3E9 in complex with the N-terminal domain of *Pv*RBP2b were obtained in 0.2 M KBr, 15% (w/v) PEG 4,000 and 0.05 M sodium citrate/citric acid pH 5.5. Crystals of the 4F7 Fab fragment alone were obtained in 20% (w/v) PEG 4,000, 0.05 M sodium citrate/citric acid pH 5.5 and 20% (v/v) 2-propanol. Crystals of 4F7 in complex with the N-terminal domain of *Pv*RBP2b were obtained in 0.2 M trisodium citrate and 20% (w/v) PEG 3,350. Crystals of

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

10B12 Fab fragment alone were obtained in 20% (w/v) PEG 4,000, 0.1 M sodium citrate/citric acid pH 5.5 and 20% (v/v) 2-propanol. Crystals of 6H1 in complex with the N-terminal domain of *Pv*RBP2b were obtained in 0.2 M $(NH_4)_2SO_4$ and 20% (w/v) PEG 3,350.

Data collection and structure solution. Crystals were cryo-protected in a reservoir solution supplemented with 20% (v/v) glycerol and flash frozen in liquid nitrogen. Diffraction data were collected at the MX2 beamline at the Australian Synchrotron Facility in Clayton, Australia at 0.9537 Å wavelength using ADSC Quantum 315r detector. Data were integrated using iMosflm[38]. Scaling and merging were performed using the program Aimless from the CCP4 package[39]. Cell content was analysed using the program Matthews[40]. Molecular replacement for the data collected for 4F7 Fab fragment alone was performed using Phaser[41] with an unrelated antibody as a model (PDB ID: 5EN2). The model obtained together with the structure of the N-terminal domain of *Pv*RBP2b (PDB ID: 5W53) were used to solve structures of 3E9, 4F7 and 6H1 Fab complexes. The crystals for 10B12 contained only the Fab fragment. Initial models were rebuilt automatically using the program AutoBuild[32] followed by a manual improvement using the program Coot[14]. The structures were refined using the program Phenix Refine[32] and included TLS (translation/libration/screw) motions that were generated using the TLSMD web server[42]. Crystallographic data collection and refinement statistics for all structures are summarized in Extended Data Table 3. Figures were prepared using either Pymol (http://www.pymol.org) or Chimera[29].

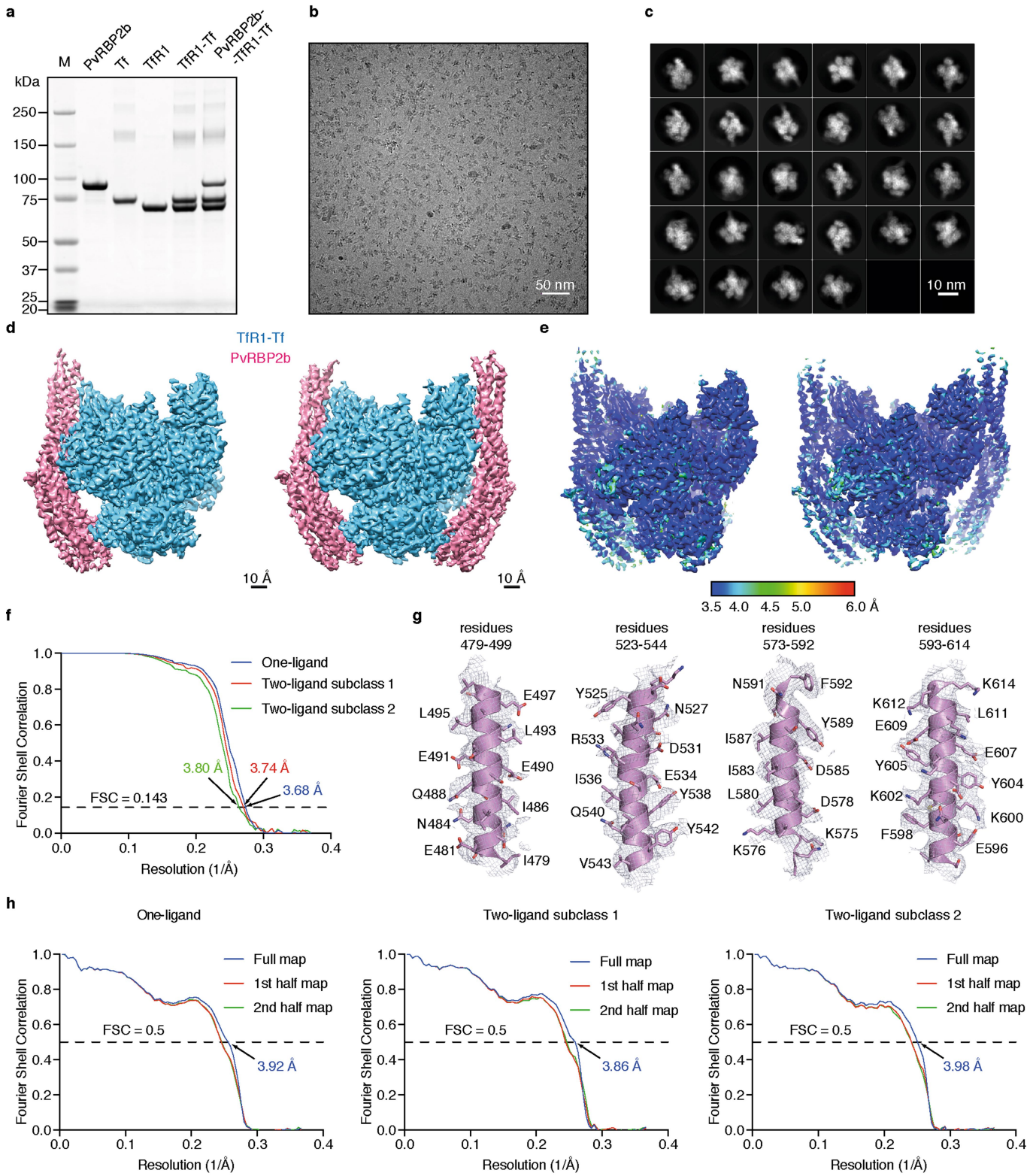
Small angle X-ray scattering. SAXS experiments were performed as described previously[2,14]. We docked the crystal structures of the N-terminal domain of *Pv*RBP2b and 10B12 Fab fragment into the SAXS envelope of the 10B12–*Pv*RBP2b complex using the program Collage from the Situs Program Package[43]. Owing to the symmetry of the SAXS envelope and the Fab fragment, we could not unambiguously assign the identity of the heavy and light chains within the model.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The cryo-EM maps have been deposited in the Electron Microscopy Data Bank with accession codes EMD-7783, EMD-7784 and EMD-7785. The coordinates of the atomic models have been deposited in the Protein Data Bank under accession codes 6D03, 6D04 and 6D05 for the *Pv*RBP2b–Tf–TfR1 one-ligand complex, two-ligand complex subclass 1 and two-ligand complex subclass 2, respectively. Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 6BPA (*Pv*RBP2b–3E9 complex), 6BPB (4F7 Fab alone), 6BPC (*Pv*RBP2b–4F7 complex), 6BPD (10B12 Fab alone) and 6BPE (*Pv*RBP2b–6H1 complex).

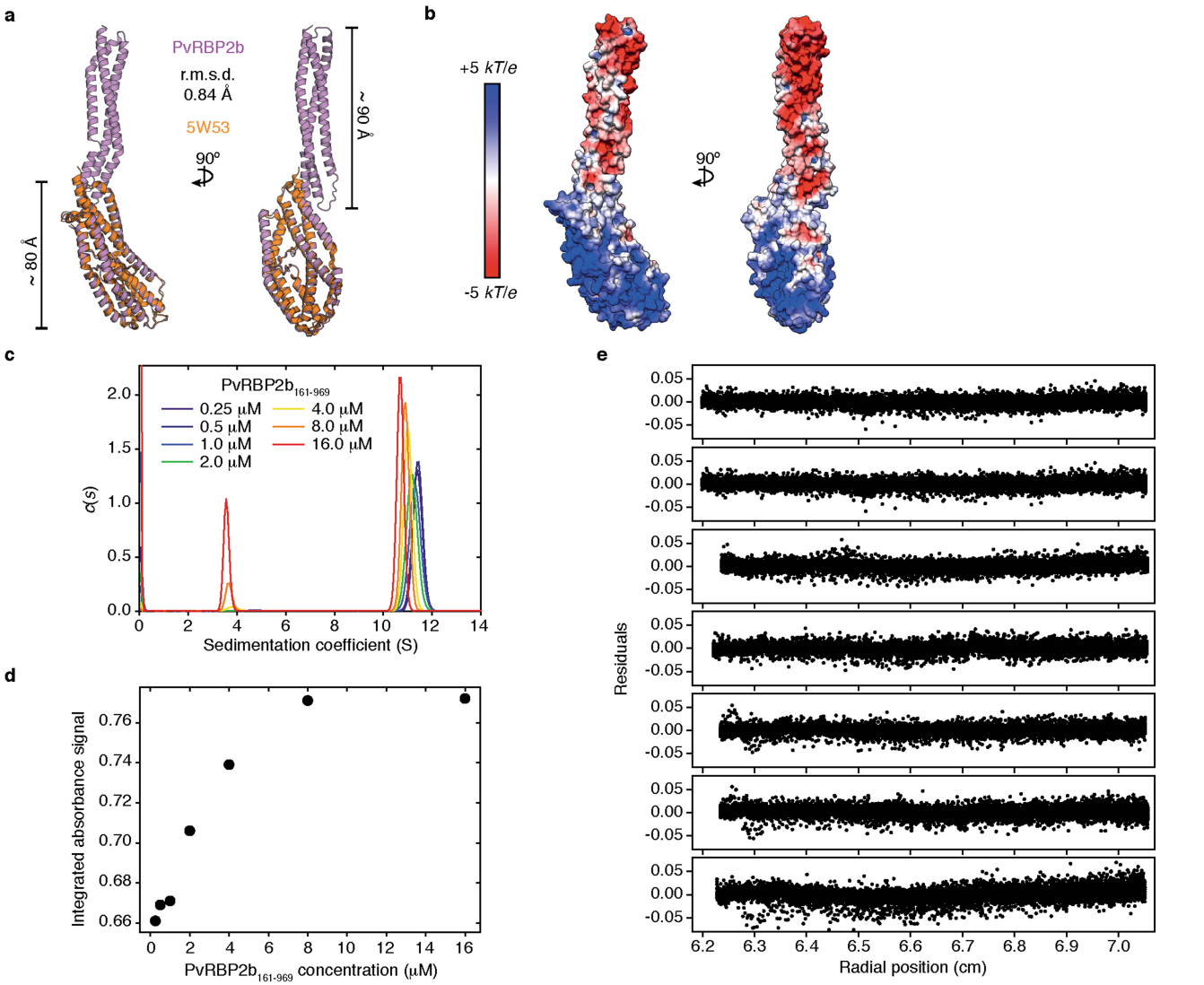
25. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
26. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
27. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
28. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
29. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
30. Yang, N., Zhang, H., Wang, M., Hao, Q. & Sun, H. Iron and bismuth bound human serum transferrin reveals a partially-opened conformation in the N-lobe. *Sci. Rep.* **2**, 999 (2012).
31. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
32. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
33. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
34. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
35. Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
36. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
37. Harding, S. E., Rowe, A. J., Horton, J. C. (eds.) *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Royal Society of Chemistry, Cambridge, 1992).
38. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D* **67**, 271–281 (2011).
39. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
40. Kantardjieff, K. A. & Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Sci.* **12**, 1865–1871 (2003).
41. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
42. Painter, J. & Merritt, E. A. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Cryst.* **39**, 109–111 (2006).
43. Birmanns, S., Rusu, M. & Wriggers, W. Using Sculptor and Situs for simultaneous assembly of atomic components into low-resolution shapes. *J. Struct. Biol.* **173**, 428–435 (2011).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



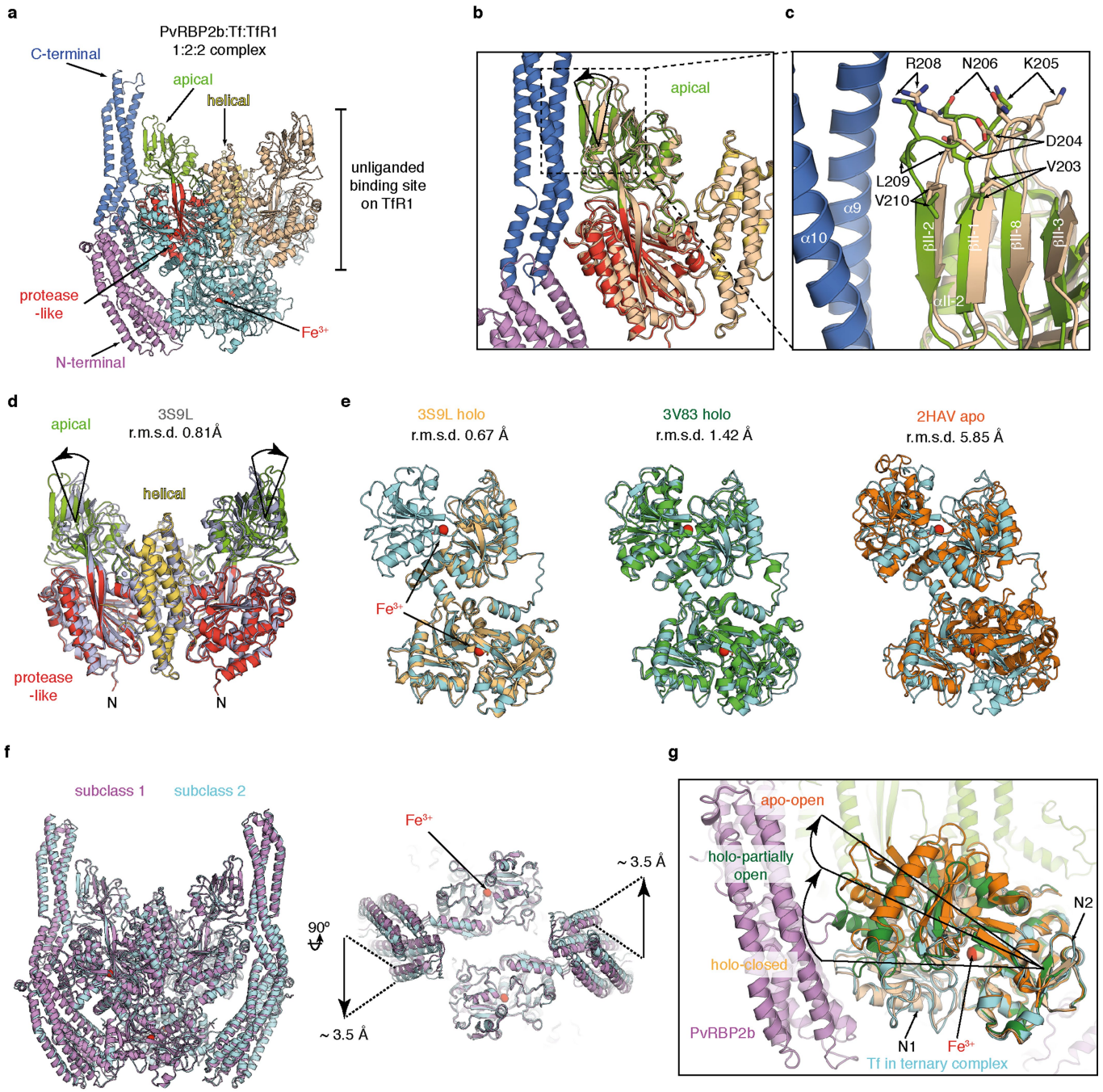
Extended Data Fig. 1 | Details of cryo-EM data collection and analysis. **a**, Coomassie-stained reducing SDS–PAGE gel of proteins used for the sample preparation. M, molecular weight marker. For gel source data, see Supplementary Fig. 1. **b**, Representative micrograph of the sample after drift correction and dose weighing. **c**, Representative 2D class averages. The 2D class averages exhibit different projections corresponding to each orientation. **d**, 3D classification effectively separated two groups of populations. Left, TfR1–Tf complex with one bound *Pv*RBP2b molecule. Right, TfR1–Tf complex with two bound *Pv*RBP2b molecules. Blue, TfR1–Tf; pink, *Pv*RBP2b. Scale bar, 10 Å. **e**, Local resolution estimation diagram of the final refined maps. Left, the one-ligand complex; right: the two-ligand complex. Resolution keys are labelled from 3.5 to 6.0 Å. **f**, Resolution estimation of the cryo-EM map. Fourier shell correlation (FSC) plot showing resolutions at 0.143 FSC (dashed line) for both complexes, determined by the gold-standard method. **g**, Representative cryo-EM density for different parts of the C-terminal domain of *Pv*RBP2b. Selected residues are also indicated. **h**, FSC curves of the final refined model versus the final cryo-EM map (full dataset, blue), of the outcome of model refinement with a half map versus the same map (red), and of the outcome of model refinement with a half map versus the other half map (green). In **a**–**c**, the experiment was performed once.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



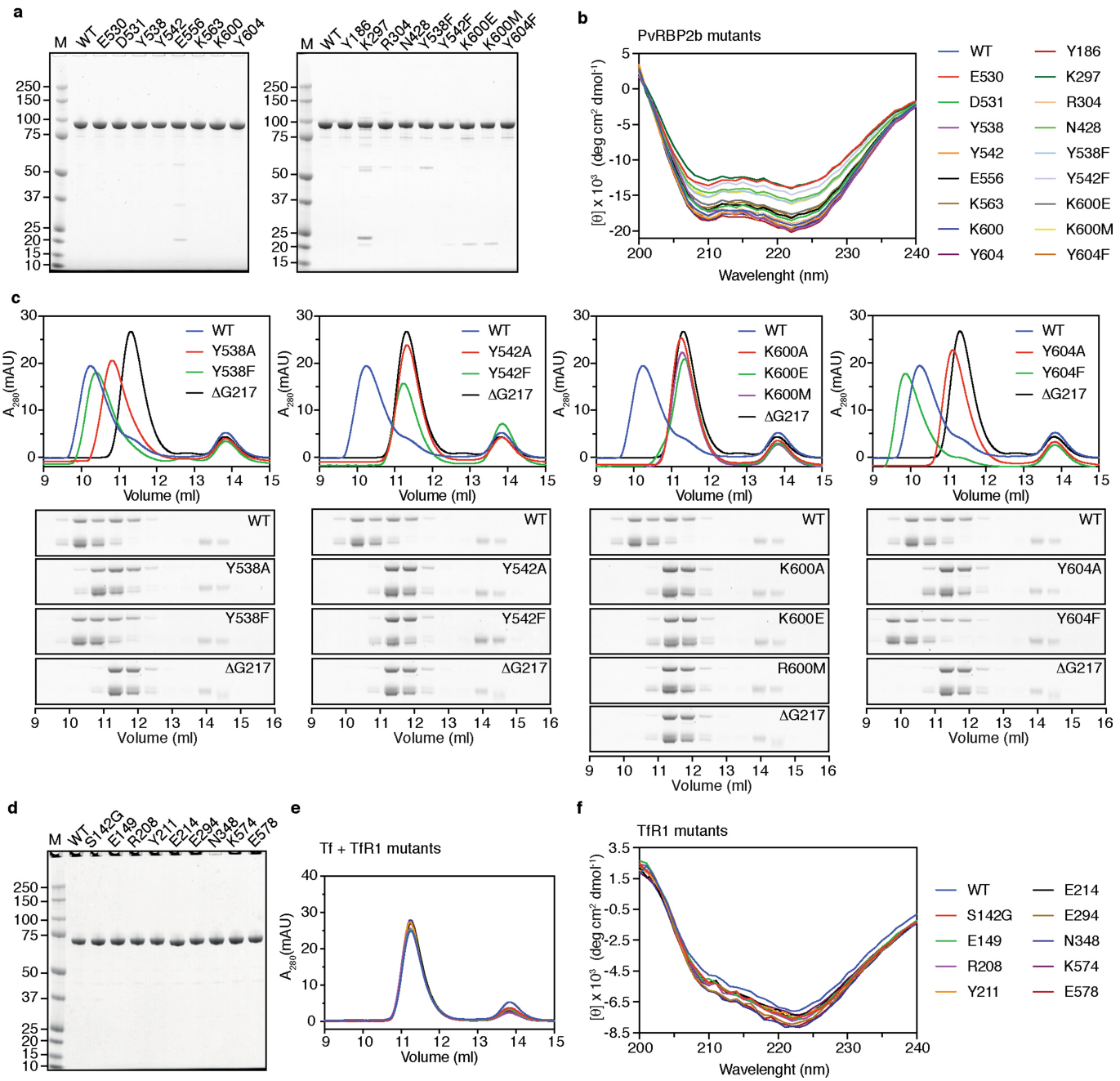
Extended Data Fig. 2 | Detailed structural analyses of *Pv*RBP2b and sedimentation velocity analysis of the formation of the ternary *Pv*RBP2b(161–969)–TfR1–Tf complex. a, Superposition of the cryo-EM structure of *Pv*RBP2b with the previously reported crystal structure of the N-terminal domain (PDB ID: 5W53). **b**, Two orthogonal views of *Pv*RBP2b shown in surface representation and coloured according to the potential on the surface of the molecule. Electrostatic surface potential was calculated using the program APBS in Chimera with the nonlinear Poisson–Boltzmann equation and contoured at $\pm$ 5 kT e^{-1}. Negatively and positively charged surface areas are coloured red and blue, respectively. **c**, Continuous sedimentation coefficient ($c(s)$) distributions for the TfR1–Tf complex (2.5 μM complex) in the presence of increasing concentrations of *Pv*RBP2b(161–969) (0.25–16 μM). The configuration and surface charge properties of the ternary complex suggest that this particle may display non-ideal sedimentation, possibly contributing to the observed shift in sedimentation coefficient. Thus, the integrated absorbance signal in the fast-sedimenting peak of the distributions was analysed as a function of *Pv*RBP2b(161–969) concentration. **d**, Here the TfR1–Tf complex was assumed to sediment as a single, stable species. These data are consistent with the binding of two molecules of *Pv*RBP2b(161–969) to the TfR1–Tf complex with macroscopic dissociation constants in the mid-nanomolar to low-micromolar range. **e**, Residuals for the best fit of the raw radial absorbance sedimentation velocity data to a $c(s)$ distribution model for 2.5 μM of TfR1–Tf in the presence of (top to bottom): 0.25 μM, 0.5 μM, 1.0 μM, 2.0 μM, 4.0 μM, 8.0 μM and 16.0 μM *Pv*RBP2b(161–969).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



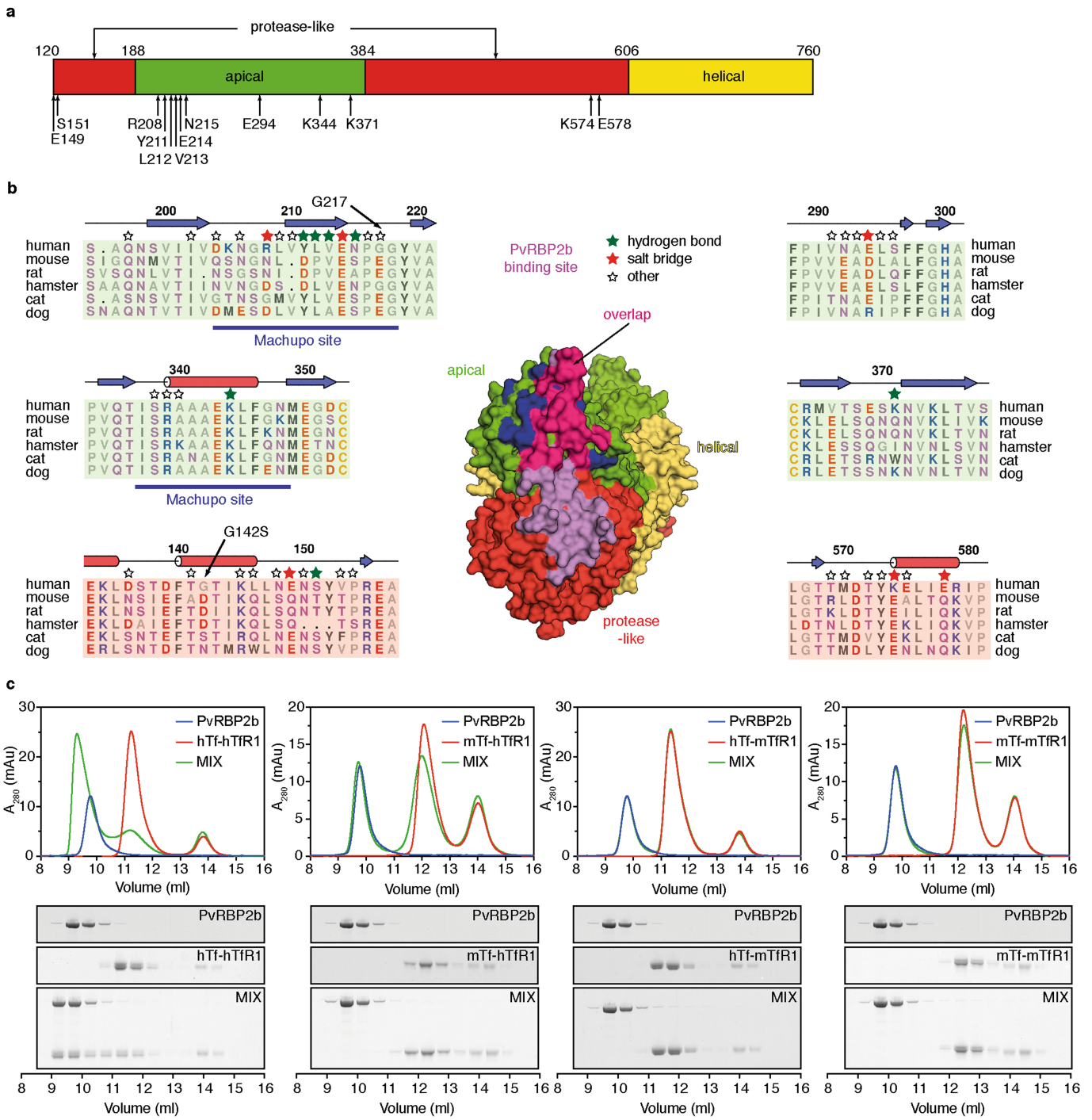
Extended Data Fig. 3 | Structure of the one-ligand complex and conformational changes induced in TfR1 upon *Pv*RBP2b binding. **a**, Overall view of the one-ligand complex. Proteins are shown in ribbon representation. The molecule of TfR1 interacting with *Pv*RBP2b is coloured according to domain organization. Protease-like domain, red; apical domain, green; helical domain, yellow. The other molecule of TfR1 that remains unliganded is coloured in wheat. Tf, cyan; Fe^{3+}, red spheres. *Pv*RBP2b N-terminal domain, violet; *Pv*RBP2b C-terminal domain, blue. **b**, Superimposition of the unliganded TfR1 molecule from the one-ligand complex on the ligand-bound TfR1 from the two-ligand complex. Colour-coded as in **a**. The movement of the apical domain is indicated schematically by a black arrow. For clarity, both Tf molecules have been omitted. **c**, Closer view of the superimposition shown in **b**. The residues that are most affected by *Pv*RBP2b binding are shown as sticks and labelled. **d**, The superposition of TfR1 in complex with Tf and *Pv*RBP2b (coloured according to domain), and TfR1 in the complex with Tf only (grey, PDB ID: 3S9L). The r.m.s.d. between the 1,090 aligned Cα atoms is 0.81 Å. The movement of the apical domain is represented schematically with black arrows. **e**, Superposition of Tf in the cryo-EM structure of *Pv*RBP2b–TfR1–Tf complex (in cyan) with the crystal structure of holo-Tf in Tf–TfR1 complex (left, wheat, PDB ID: 3S9L) or with the crystal structure of holo-Tf (middle, green, PDB ID: 3V83) or with the structure of apo-Tf (right, orange, PDB ID: 2HAV). Fe^{3+}, red spheres. R.m.s.d. values are indicated above each superimposition. **f**, Two orthogonal views of the superposition between two subclasses for the two-ligand complex. Molecules are shown in cartoon representation and coloured in violet and cyan for subclasses 1 and 2, respectively. The most important difference between the two subclasses is the movement of the N-terminal domain of *Pv*RBP2b, indicated by the black arrows. **g**, Close view of the *Pv*RBP2b–Tf interaction site. The N2 subdomain of Tf in the cryo-EM structure (cyan) is superimposed in the crystal structures of holo-Tf in closed conformation (wheat), holo-Tf in partially open conformation (green) and apo-Tf in open conformation (orange). The movement of the N1 subdomain is highlighted schematically with black arrows.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



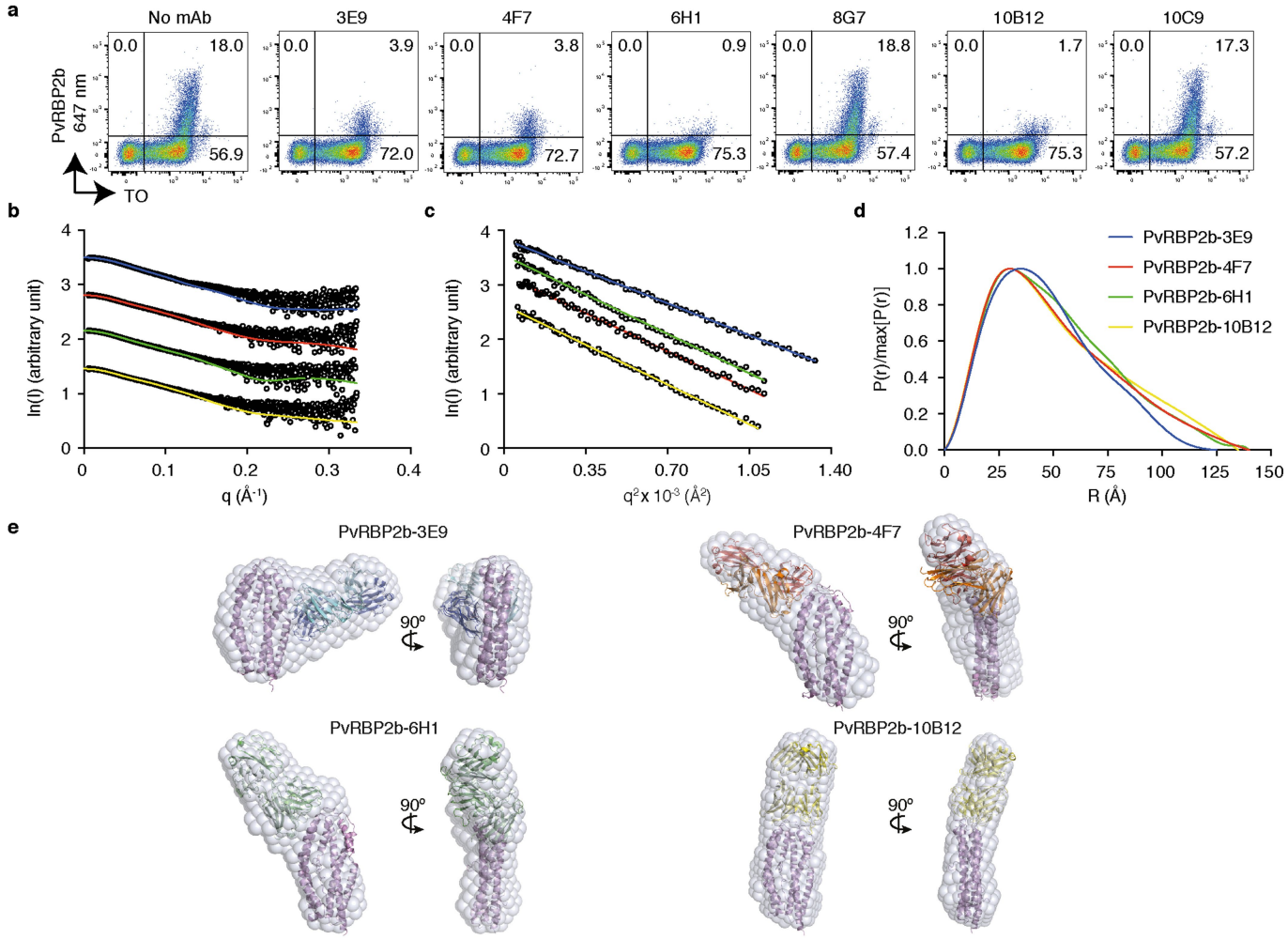
Extended Data Fig. 4 | Site-directed mutagenesis of *Pv*RBP2b and TfR1. **a**, SDS–PAGE of purified *Pv*RBP2b mutant recombinant proteins. Two micrograms of each protein were loaded onto a 4–12% NuPAGE gradient gel under reducing conditions and stained with Coomassie blue. **b**, Circular dichroism spectra of recombinant *Pv*RBP2b mutants. **c**, Analytical SEC (left to right panel) of complex formation between TfR1–Tf and *Pv*RBP2b with mutant residues Y538, Y542, K600 or Y604. SDS–PAGE of fractions from the SEC analyses. **d**, SDS–PAGE gel of purified TfR1 mutant recombinant proteins. Two micrograms of each protein were loaded onto a 4–12% NuPAGE gradient gel under reducing conditions and stained with Coomassie blue. **e**, Analytical SEC of TfR1 mutants in complex with Tf. **f**, Circular dichroism spectra of recombinant TfR1 mutants. All experiments were performed once. For gel source data, see Supplementary Fig. 1.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Fig. 5 | Most TfR1 residues that interact with *Pv*RBP2b are specific to human TfR1. a, Schematic representation of TfR1 sequence coloured according to domain. Apical domain, green; protease-like domain, red; helical domain, yellow. Domain boundaries are indicated above the schematic. The most important residues forming either hydrogen bonds or salt bridges with *Pv*RBP2b are indicated below with black arrows. **b**, Comparison between Machupo virus GP1 and *Pv*RBP2b binding sites. Human TfR1 is shown in the centre in surface representation and coloured according to the domain organization. The residues that interact with *Pv*RBP2b are shown in violet and those that interact with Machupo virus GP1 are dark blue. Residues overlapping the two sites are shown in magenta. The surrounding panels show fragments of the alignment between TfR1 sequences from different species (human, mouse, rat, hamster, cat and dog). Residues are coloured according to their chemical properties (acidic, red; basic, blue; polar, magenta; cysteine, yellow; hydrophobic, grey scale with intensity proportional to hydrophobicity). The fragments of the alignment corresponding to the apical and protease-like domains of TfR1 are additionally shaded in green and red, respectively. The residues interacting with *Pv*RBP2b are indicated with stars above the alignment, which are coloured according to the type of interaction (hydrogen bonds, green; salt bridges, red; others, empty stars). The Machupo virus GP1-interaction site is indicated with a dark blue line. The position of the G217 deletion and the G142S polymorphism are indicated with black arrows above the alignment. **c**, Analytical SEC results showing *Pv*RBP2b(161–1454) interaction with mouse Tf and TfR1. For gel source data, see Supplementary Fig. 1. Top, elution profiles for the separate components and for the combined components. Bottom, SDS–PAGE of SEC fractions. Left, human Tf and human TfR1; middle left, mouse Tf and human TfR1; middle right, human Tf and mouse TfR1; right, mouse Tf and mouse TfR1. The same representative SDS–PAGE gel is used for *Pv*RBP2b protein SEC analyses. All analytical SEC experiments were performed once.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



f

	PvRBP2b-3E9	PvRBP2b-4F7	PvRBP2b-6H1	PvRBP2b-10B12
Data collection				
Instrument	Australian Synchrotron SAXS/WAXS beamline			
Beam geometry	120 μm point source			
Wavelength (Å)	1.033			
q range (Å^{-1})	0.005-0.33			
Exposure time (s)	2			
Protein concentration	~5.0 mg/ml protein via in-line gel filtration chromatography			
Temperature (K)	289			
Structural parameters				
$I(0)$ (cm^{-1}) [from $P(r)$]	0.024 ± 0.002	0.020 ± 0.002	0.022 ± 0.002	0.018 ± 0.002
R_g (Å) [from $P(r)$]	36.80 ± 0.23	40.62 ± 0.34	40.57 ± 0.43	41.75 ± 0.30
$I(0)$ (cm^{-1}) [from Guinier]	0.024 ± 0.001	0.020 ± 0.001	0.022 ± 0.001	0.018 ± 0.001
R_g (Å) [from Guinier]	35.50 ± 0.35	38.80 ± 0.55	38.80 ± 0.51	39.40 ± 0.60
D_{max} (Å)	125	140	140	135
Porod volume estimate (Å^3)	112,088	106,141	110,075	109,816
NSD	0.781 ± 0.083	0.749 ± 0.030	0.815 ± 0.056	0.778 ± 0.038
χ	0.65	0.54	0.50	0.50
Software employed				
Primary data reduction	ScatterBrain (Australian Synchrotron)			
Data processing	PRIMUS, GNOM			
Ab initio analysis	DAMMIF, DAMSEL, DAMSUP			
Validation and averaging	DAMAVER, DAMFILT			
Rigid-body modeling	CRYSOL	CRYSOL	CRYSOL	COLLAGE
Computation of model intensities	CRYSOL			
3D graphics representation	MacPyMOL			

Extended Data Fig. 6 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Fig. 6 | Reticulocyte-binding and SAXS data analyses for the N-terminal domain of *Pv*RBP2b in complexes with four different inhibitory Fab fragments. **a**, *Pv*RBP2b(161–1454) binding in the presence or absence of anti-*Pv*RBP2b monoclonal antibodies (3E9, 4F7, 6H1, 8G7 and 10B12) or anti-PfRh4 monoclonal antibodies (10C9), analysed by flow cytometry. Dot plots of *Pv*RBP2b(161–1454) binding (*y* axis) to reticulocytes stained with thiazole orange (TO, *x* axis). This experiment was repeated independently four times with similar results. **b**, Arbitrarily offset scattering intensity profiles for the *Pv*RBP2b–Fab complexes. The background-subtracted SAXS data are shown as black open circles representing natural logarithm of mean intensity $\ln I(q)$ as a function of momentum transfer q in $Å^{-1}$. a.u., arbitrary unit. The theoretical scattering profiles (solid lines: *Pv*RBP2b–3E9, blue; *Pv*RBP2b–4F7, red; *Pv*RBP2b–6H1, green; *Pv*RBP2b–10B12, yellow) calculated from the crystal structures were fitted to the experimental scattering data using CRYSOL. **c**, Guinier plots for $qR_g \leq 1.3$ (qR_g is the product of momentum transfer (q) and the radius of gyration (R_g)), showing that neither high-molecular-mass aggregates nor inter-particle interference contribute measurably to scattering and the data are of high quality (colours as in **b**). **d**, Pair-wise inter-atomic distance distribution function, $P(r)$ (colours as in **b**). **e**, Two orthogonal views of ab initio bead models represented as grey spheres, superimposed with the crystal structure of the corresponding antibody complexes. For the *Pv*RBP2b–10B12 complex, COLLAGE was used to rigid-body fit two crystal structures, the N-terminal domain of *Pv*RBP2b (PDB ID: 5W53) and the structure of the 10B12 Fab fragment alone. **f**, Table summarizing SAXS data collection and analysis. The radius of gyration and initial scattering intensity $I(0)$ were approximated using the Guinier equation with PRIMUS. The radius of gyration and maximum particle dimension Dmax were calculated from the $P(r)$ analysis using program GNOM. The normalized spatial discrepancy parameter (NSD) indicates the similarity between 20 independently generated ab initio models. The χ parameter provides the assessment of the fit of the average model to the experimental data calculated using program CRYSOL. The structural parameters and their associated errors are those derived by least-squares curve-fitting analysis of the scattering data as formulated within the listed software packages.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	One-ligand complex (EMD-7783) (PDB 6D03)	Two-ligand complex subclass 1 (EMD-7784) (PDB 6D04)	Two-ligand complex subclass 2 (EMD-7785) (PDB 6D05)
Data collection and processing			
Magnification (calibrated)	37,037	37,037	37,037
Voltage (kV)	300	300	300
Electron exposure (e–/Å^2)	80	80	80
Defocus range (μm)	-0.8 to -2.2	-0.8 to -2.2	-0.8 to -2.2
Pixel size (Å)	1.35	1.35	1.35
Number of particle images	637,649	287,253	302,858
Refinement			
Initial model used (PDB code)	1SUV	1SUV	1SUV
Initial model resolution (Å)	40	40	40
Map resolution (Å)	3.68	3.74	3.80
FSC threshold	0.143	0.143	0.143
Map sharpening *B* factor (Å^2)	-197.61	-181.95	-185.62
Symmetry imposed	C1	C2	C2
Model composition			
Non-hydrogen atoms	24,844	28,748	28,748
Protein residues	3,106	3,572	3,572
Ligands	246	246	246
B factors (Å^2)			
Protein	41.82	68.93	76.06
Ligand	67.37	91.47	98.11
R.m.s. deviations			
Bond lengths (Å)	0.010	0.009	0.009
Bond angles (°)	1.34	1.24	1.28
Validation			
MolProbity score	1.63	1.58	1.64
Clashscore	4.31	4.28	4.42
Poor rotamers (%)	1.15	1.15	1.34
EMRinger score	3.21	2.28	1.81
Ramachandran plot			
Favored (%)	94.6	95.4	95.4
Allowed (%)	5.3	4.6	4.6
Disallowed (%)	0.1	-	-

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 2 | Summary of interactions between *Pv*RBP2b and TfR1–Tf

TfR1	Group	Location	PvRBP2b	Group	Location	Distance (Å)
			Hydrogen bonds			
Ser 151	OG	loop αI-2-αI-3	Lys 363	NZ	loop α5-α6	4.0
Arg 208	NH2	loop βII-1-βII-2	Asn 527	O	helix α9	2.3
Tyr 211	OH	strand βII-2	Ser 541	OG	helix α9	2.7
Leu 212	O	strand βII-2	Tyr 542	OH	helix α9	3.9
Leu 212	N	strand βII-2	Tyr 542	OH	helix α9	3.4
Val 213	O	strand βII-2	Tyr 604	OH	helix α10	2.7
Asn 215	ND2	loop βII-2-βII-3	Asp 603	OD2	helix α10	3.4
Lys 344	NZ	helix αII-2	Tyr 604	OH	helix α10	3.2
Lys 371	NZ	loop αII-2-βII-8	Tyr 538	OH	helix α9	4.0
Lys 574	NZ	helix αI-9	Ser 586	OG	helix α10	3.9
Lys 574	NZ	helix αI-9	Ser 590	OG	helix α10	3.4
			Salt bridges			
Glu 149	OE2	loop αI-2-αI-3	Arg 359	NH1	helix α5	3.0
Glu 149	OE2	loop αI-2-αI-3	Arg 359	NH2	helix α5	2.9
Arg 208	NE	loop βII-1-βII-2	Asp 531	OD1	helix α9	3.4
Arg 208	NE	loop βII-1-βII-2	Asp 531	OD2	helix α9	3.7
Arg 208	NH1	loop βII-1-βII-2	Asp 531	OD1	helix α9	3.1
Arg 208	NH1	loop βII-1-βII-2	Glu 530	OE1	helix α9	3.4
Arg 208	NH2	loop βII-1-βII-2	Asp 531	OD1	helix α9	2.2
Arg 208	NH2	loop βII-1-βII-2	Asp 531	OD2	helix α9	2.4
Glu 214	OE2	strand βII-2	Lys 600	NZ	helix α10	3.8
Glu 294	OE2	loop βII-6-βII-7	Lys 600	NZ	helix α10	2.4
Lys 574	NZ	helix αI-9	Glu 556	OE1	helix α9	3.8
Glu 578	OE2	helix αI-9	Lys 563	NZ	helix α9	3.6

Tf	Group	Location	PvRBP2b	Group	Location	Distance (Å)
			Hydrogen bonds			
Ser 28	O	loop α1-β2	Arg 304	NH1	helix α3	3.5
Val 29	O	loop α1-β2	Lys 297	NZ	helix α3	4.0
Ser 32	OG	loop α1-β2	Asn 428	ND2	helix α7	3.7
Glu 265	OE2	helix α12	Tyr 186	OH	loop β1-β2	3.8

			Other interfacing residues (TfR1)			
Asp 139	Ser 142	Lys 145	Leu 146	Asn 148	Asn 150	Val 153
Pro 154	Gln 197	Ile 202	Asp 204	Asn 206	Leu 209	Val 210
Pro 216	Gly 217	Val 291	Asn 292	Ala 293	Leu 295	Ser 296
Ser 338	Arg 339	Ala 340	Gly 413	Pro 414	Thr 569	Met 570
Thr 572	Tyr 573	Glu 575				

			Other interfacing residues (Tf)			
Val 1	Pro 2	Lys 4	Asp 24	Ile 30	Pro 31	Asp 33
Gly 34	Pro 35	Gln 269				

			Other interfacing residues (PvRBP2b-TfR1)			
Asn 349	Asp 356	Lys 360	Glu 534	Ser 545	Ser 548	Ala 549
Arg 551	Tyr 552	Leu 555	Ser 559	Val 560	Asp 585	Ser 588
Tyr 589	Phe 592	Ser 593	Glu 596	Asn 597	Gln 599	Glu 607
Leu 611	Lys 614					

			Other interfacing residues (PvRBP2b-Tf)			
Asn 185	Tyr 187	Asn 301	Lys 394	His 397	Gln 401	Asn 417
Glu 421	Phe 424	Asn 425	Ile 432			

The distance measurements are based on molecules B, C and F in two-ligand complex subclass 1.

The distance measurements are based on molecules B, C and F in two-ligand complex subclass 1.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 3 | Data collection and refinement statistics for *Pv*RBP2b complexes with Fab fragments

	PvRBP2b-3E9 (PDB 6BPA)	4F7 Fab alone (PDB 6BPB)	PvRBP2b-4F7 (PDB 6BPC)	10B12 Fab alone (PDB 6BPD)	PvRBP2b-6H1 (PDB 6BPE)
Data collection					
Space group	P 2 2_1 2_1	P 2 2_1 2_1	P 1	C 2	P 2_1
Cell dimensions					
a, *b*, *c* (Å)	59.07, 177.10, 178.62	40.72, 63.34, 182.64	70.09, 70.31, 140.36	118.54, 42.66, 105.03	106.60, 123.48, 143.28
α, β, γ(°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	104.18, 100.34, 98.41	90.00, 122.03, 90.00	90.00, 105.25, 90.00
Resolution (Å)	59.54-2.53 (2.59-2.53)[a]	43.89-1.87 (1.91-1.87)	66.55-2.66 (2.72-2.66)	50.25-2.32 (2.40-2.32)	75.49-3.34 (3.44-3.34)
R_{merge}	0.168 (1.128)	0.099 (1.028)	0.121 (0.580)	0.112 (0.703)	0.126 (0.785)
$I/\sigma(I)$	9.9 (1.9)	7.7 (1.6)	6.0 (1.9)	8.5 (1.6)	5.5 (1.2)
$CC_{1/2}$	0.993 (0.552)	0.994 (0.545)	0.982 (0.580)	0.992 (0.539)	0.992 (0.612)
Completeness (%)	99.8 (99.3)	86.9 (85.7)	94.3 (94.3)	99.5 (98.8)	99.5 (99.2)
Redundancy	6.7 (6.4)	5.6 (6.0)	3.6 (3.5)	3.5 (3.5)	3.3 (3.2)
Refinement					
Resolution (Å)	56.05-2.53	43.89-1.87	66.31-2.66	44.53-2.32	73.75-3.34
No. reflections	63,452	34,753	67,840	18,236	51,741
R_{work} / R_{free} (%)	19.43 / 22.43	20.62 / 21.89	19.73 / 21.48	21.08 / 26.19	31.75 / 36.67
No. atoms					
Protein	11,516	3,239	11,425	3,056	17,777
Ligand/ion	1 (Br)	-	-	-	-
Water	542	347	620	178	-
B factors					
Protein	44.33	38.83	48.94	41.23	86.88
Ligand/ion	67.28	-	-	-	-
Water	38.38	43.00	46.57	35.48	-
R.m.s. deviations					
Bond lengths (Å)	0.004	0.004	0.005	0.006	0.004
Bond angles (°)	0.640	0.651	0.675	0.994	0.720
Validation					
MolProbity score	1.52	1.42	1.87	1.90	2.11
Clashscore	6.52	7.72	14.68	9.99	20.40
Poor rotamers (%)	0.15	-	0.93	-	-
Ramachandran plot					
Favored (%)	97.1	98.6	96.8	94.5	95.6
Allowed (%)	2.4	1.4	2.8	3.9	4.0
Disallowed (%)	0.5	-	0.4	1.6	0.4

X-ray diffraction data were collected on single crystals.
[a]Values in parentheses are for highest-resolution shell.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 4 | Summary of interactions between *Pv*RBP2b and Fab fragments of monoclonal antibodies 3E9, 4F7 and 6H1

PvRBP2b and 3E9 Fab fragment based on the crystal structure PDB 6BPA

PvRBP2b	Group	Location	3E9 V_H	Group	Distance (Å)
Hydrogen bonds					
Lys 333	NZ	α4	His 51	O	2.7
Lys 396	O	α6	Ser 122	OG	3.3
Lys 396	NZ	α6	Ser 123	OG	3.6
Asp 400	OD1	α6	Ser 120	OG	2.7
Asp 400	OD1	α6	Ser 122	OG	2.9
Asp 400	OD1	α6	Ser 123	OG	2.6
Asp 407	O	α6	Glu 20	N	3.7
Asp 407	OD2	α6	Tyr 128	OH	3.3
Salt bridges					
Lys 333	NZ	α4	Asp 71	OD1	3.0
Lys 333	NZ	α4	Asp 71	OD2	3.7
Glu 338	OE1	α5	Arg 69	NH1	3.2
Glu 338	OE2	α5	Arg 69	NH1	3.5
Glu 338	OE2	α5	Arg 69	NH2	3.0
Asp 407	OD1	α6	Arg 117	NH1	3.1
Asp 407	OD1	α6	Arg 117	NH2	3.8
Asp 407	OD2	α6	Arg 117	NH1	3.4
Asp 407	OD2	α6	Arg 117	NH2	2.7

PvRBP2b	Group	Location	3E9 V_L	Group	Distance (Å)
Hydrogen bonds					
Lys 345	NZ	α5	Asp 114	O	3.7
Salt bridges					
Lys 345	NZ	α5	Asp 54	OD1	2.8
Lys 345	NZ	α5	Asp 114	OD1	2.8

Other PvRBP2b interfacing residues (3E9 V_H)

Lys 326	Leu 329	Val 330	Phe 332	Met 335	Tyr 337
Ile 399	Thr 403	Ser 406	Gly 408	Val 409	Lys 410

Other PvRBP2b interfacing residues (3E9 V_L)

Glu 338	Asp 341	Thr 342	Lys 348	Asn 349	Lys 396

PvRBP2b and 4F7 Fab fragment based on the crystal structure PDB 6BPC

PvRBP2b	Group	Location	4F7 V_H	Group	Location
Hydrogen bonds					
Arg 217	NH1	α2	Thr 119	O	2.9
Arg 217	NH1	α2	Val 120	O	3.2
Asp 220	OD2	α2	Tyr 52	OH	3.1
Asp 220	OD2	α2	Asn 71	ND2	2.5
Lys 223	NZ	α2	Ser 75	OG	3.3
Lys 225	NZ	α2	Asp 50	O	2.9
Glu 302	OE2	α3	Tyr 72	OH	3.2
Lys 309	NZ	α3	Val 121	OH	3.1
Salt bridges					
Arg 217	NH1	α2	Glu 118	OE2	3.2
Arg 217	NH2	α2	Glu 118	OE2	3.0
Lys 223	NZ	α2	Asp 73	OD1	3.3
Lys 223	NZ	α2	Asp 73	OD2	2.4
Lys 298	NZ	α3	Asp 50	OD2	3.8

PvRBP2b	Group	Location	4F7 V_L	Group	Location
Hydrogen bonds					
Arg 217	NH2	α2	Tyr 117	OH	3.4
Arg 217	NH2	α2	Tyr 119	OH	3.5
Asp 305	OD1	α3	Tyr 73	OH	3.9
Lys 309	NZ	α3	Tyr 114	O	2.3
Lys 309	NZ	α3	Asn 115	O	3.7
Lys 311	NZ	α3-α4	Asn 52	O	3.7
Lys 416	NZ	α6-α7	Thr 55	OG1	3.6

Other PvRBP2b interfacing residues (4F7 V_H)

Ile 221	Thr 224	Gln 295	Asp 305	Leu 306	Glu 308

Other PvRBP2b interfacing residues (4F7 V_L)

Glu 214	Glu 308	Leu 310	Ser 313	Asn 417

PvRBP2b and 6H1 Fab fragment based on the crystal structure PDB 6BPE

PvRBP2b	Group	Location	6H1 V_H	Group	Location
Hydrogen bonds					
Arg 217	NH1	α2	Val 120	O	3.5
Arg 217	NH2	α2	Tyr 72	OH	2.2
Asp 220	NZ	α2	Ser 123	OG	3.8
Glu 302	OE2	α3	Tyr 72	OH	3.2
Lys 309	NZ	α3	Val 120	O	3.3
Salt bridges					
Lys 225	NZ	α2	Asp 50	OD1	3.0
Lys 298	NZ	α3	Asp 50	OD2	3.9
Lys 309	NZ	α3	Asp 122	OD1	2.3
Lys 309	NZ	α3	Asp 122	OD2	3.8

Other PvRBP2b interfacing residues (6H1 V_H)

Ile 221	Lys 223	Thr 224	Gln 295	Asp 305	Leu 306
Glu 308					

PvRBP2b	Group	Location	6H1 V_L	Group	Location
Hydrogen bonds					
Lys 309	O	α3	Asn 56	ND2	3.4
Lys 311	NZ	α3-α4	Ser 117	OG	3.6
Cys 312	O	α3-α4	Thr 55	OG1	3.6
Ser 313	OG	α3-α4	Ser 91	OG	2.9
Lys 416	NZ	α6-α7	Thr 55	OG1	3.2
Salt bridges					
Lys 311	NZ	α3-α4	Asp 116	OD1	2.9
Lys 311	NZ	α3-α4	Asp 116	OD2	2.5

Other PvRBP2b interfacing residues (6H1 V_L)

Glu 214	Glu 308	Leu 310	Asn 417

The distance measurements are based on molecules A, B and C.

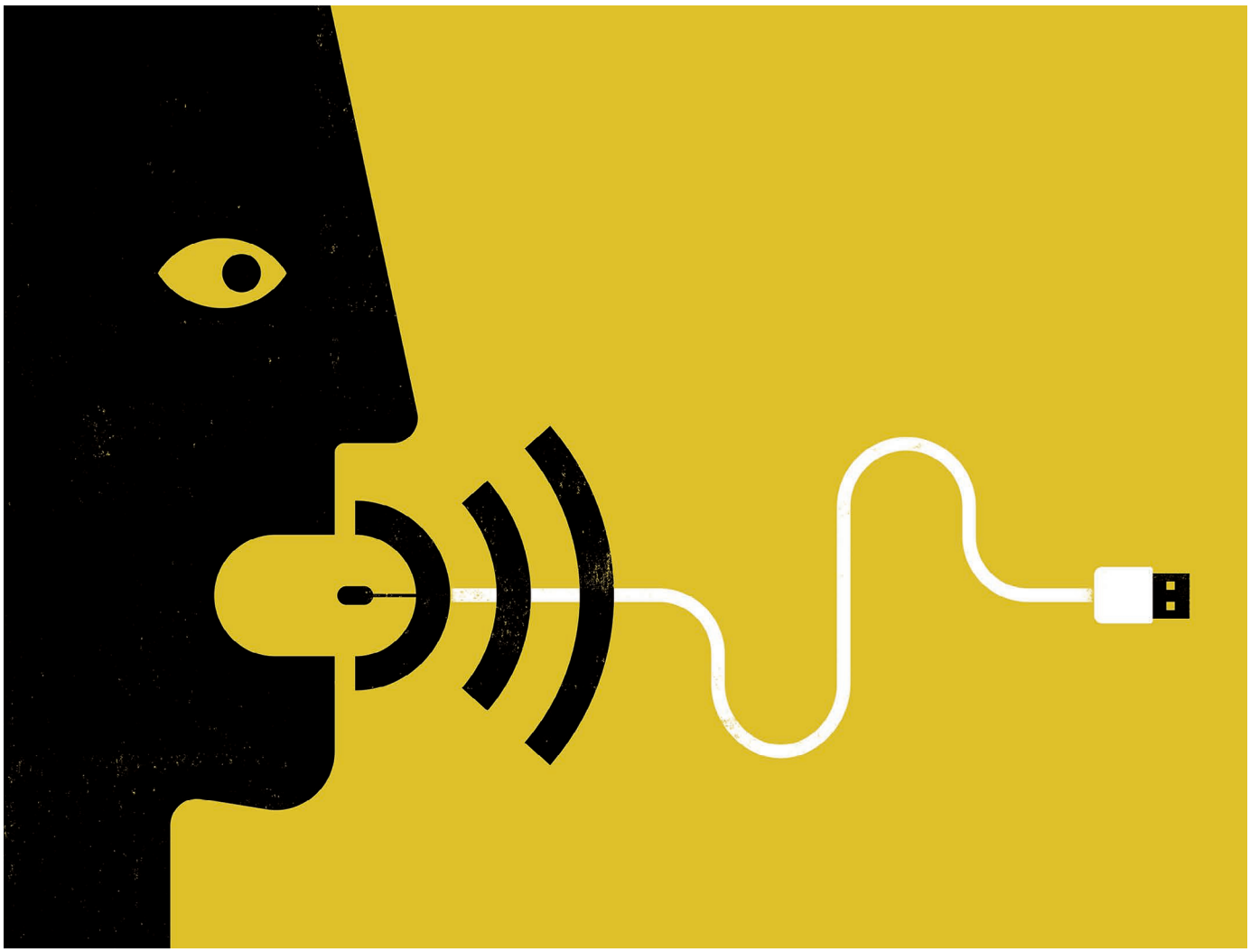
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

TOOLBOX

WRITING CODE OUT LOUD

Programmers turn to voice-command tools to give their hands a rest.

ILLUSTRATION BY THE PROJECT TWINS



BY ANNA NOWOGRODZKI

Debilitating hand pain is always bad news, but Harold Pimentel's was especially unwelcome. As a computational-biology PhD student, his work involved constant typing — and he was born with only one arm. "My adviser jokingly said, 'Can't you do this by voice?'" he recalls. Three years later, as a computational-genomics postdoc at Stanford University in California, he does just that.

Pimentel had cubital tunnel syndrome caused by repetitive strain injury (RSI). The syndrome occurs when the ulnar nerve, which travels down the outer edge of the arm, becomes pinched at the elbow, causing numbness, pain and loss of fine motor control in the hands and fingers. RSI can derail the careers of computational biologists and other scientists who code. Now, a small but growing community has developed a workaround: coding by voice command. It takes at least a month of difficult, sometimes frustrating, training to get set up, but coding by voice helps these programmers to keep doing their jobs or continue their studies. And they say that there are unexpected advantages.

YOUTUBE INSPIRATION

Voice coding underlies a wide variety of science — any researcher who writes code could use it. Matthew Solomonson, a software engineer at the Broad Institute of MIT and ▶

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Harvard in Cambridge, Massachusetts, uses it to build web applications such as the Genome Aggregation Database (gnomAD), which is used to explore genomic data. "These applications share data from some of the largest sequencing studies in the world," he says.

Naomi Saphra, a PhD student in language cognition and computing at the University of Edinburgh, UK, has small-fibre neuropathy, the cause of which is unknown. It is a permanent condition of the nerves that connect the brain to the hands and feet, and causes the nerves to transmit pain in response to sensations that are not usually painful. She uses her code to explore the training process of neural language models. And Pimentel studies how retention of the non-coding sections of RNA determines tissue specificity and disease susceptibility.

Similarly to many other scientists, Pimentel and Saphra realized that voice coding was possible thanks to a video of Tavis Rudd, now director of technology at web-development firm Unbounce, demonstrating the process live at the PyCon 2013 conference for users of the Python programming language.

In that video, Rudd describes his struggle with RSI — the result of constant coding in the emacs text editor, a condition he calls 'emacs pinkie' — and his strategy for overcoming it. He developed a solution through months of painstaking work, and he calls it a "three-headed beast" because it runs three operating systems from one laptop. In front of the crowd at the conference, he used his method to dictate code instructing his laptop to read aloud a snippet of Monty Python's Dead Parrot sketch.

"It was pretty inspirational," Pimentel says. But the process also was "super buggy", Pimentel reports, and lacked an active user community to help fix the glitches. He began looking for alternatives, as did Saphra.

Coding by voice command requires two kinds of software: a speech-recognition engine and a platform for voice coding. Dragon from Nuance, a speech-recognition software developer in Burlington, Massachusetts, is an advanced engine and is widely used for programming by voice, with Windows and Mac versions available. Windows also has its own built-in speech recognition system. On the platform side, VoiceCode by Ben Meyer and Talon by Ryan Hileman (both are for Mac OS only) are popular.

Two other platforms for voice programming are Caster and Aenea, the latter of which runs on Linux. Both are free and open source, and enable voice-programming functionality in Dragonfly, which is an open-source Python framework that links actions with voice commands detected by a speech-recognition engine. Saphra tried Dragonfly, but found that setup required more use of her hands than she could tolerate.

All of these platforms for voice command work independently of coding language and text editor, and so can also be used for tasks outside programming. Pimentel, for instance, uses voice recognition to write e-mails, which he finds easier, faster and more natural than typing.

STACCATO BURSTS

To the untrained ear, coding by voice command sounds like staccato bursts of a secret language. Rudd's video is full of terms like 'slap' (hit return), 'sup' (search up) and 'mara' (mark paragraph).

Unlike virtual personal assistants such as Apple's Siri or Google's Alexa, VoiceCode and Talon don't do natural-language processing, so spoken instructions have to precisely match the commands that the system already knows. But both platforms use continuous command recognition, so users needn't pause between commands, as Siri and Alexa require.

VoiceCode commands typically use words not in the English language, because if you use an English word as a command, such as 'return', it means you can never type out that word. By contrast, Talon, Aenea and Caster feature dynamic grammar, a tool that constantly updates which words the software can recognize on the basis of which applications are open. This means users can give English words as commands without causing confusion.

In addition to voice recognition, Talon can also replace a computer mouse with eye tracking, which requires a Tobii 4c eye tracker (US$150). Other eye-mousing systems generally require both the eye tracker and head-tracking hardware, such as the TrackIR from NaturalPoint. "I want to make fully hands-free use of every part of a desktop computer a thing," says Hileman. Other mouse replacements also exist; Pimentel uses one called SmartNav.

Voice command requires at least a decent headset or microphone. Many users choose a unidirectional microphone so that others can talk to them while they are dictating code. One such mic, a cardioid mic, requires special equipment to supply power, and hardware costs can reach $400, says Pimentel.

The software can cost several hundred dollars too. The speech-recognition engine Dragon Professional costs $300, as does VoiceCode. Caster and Aenea are free and open source. Talon is available for free, but requires a separate speech-recognition engine. A beta version of Talon that includes a built-in speech-recognition engine is currently available to Hileman's Patreon supporters for $15 per month. "This kind of tech needs to be as free and widespread as possible, because I feel like we're sitting on an RSI epidemic in progress and nobody is talking about it," says Hileman. But, he adds, a "huge goal" of his is to convince lots of people who are not yet experiencing problems".

IT TAKES A VILLAGE

Whether or not users have RSI, it can be difficult and frustrating to start programming by voice. It took a month and a half for Pimentel to get up to speed, he says, and there were days when he was ready to throw in the towel. He printed out 40 pages of commands and forced himself to look at them until he learnt them. Saphra needed two months of coding, a little every day, before she felt that it was a "perfectly enjoyable experience and I could see myself doing this for a living".

After the initial learning curve, users often create custom prompts for commonly used commands as the need arises. Saphra has written prompts for creating fractions in the mathematical-typesetting system LaTeX.

Users often share their configuration files and set-up details on sites such as GitHub and Slack. VoiceCode's Slack channel has more than 250 users, of whom perhaps 40 are active, Pimentel estimates; the Talon Slack has more than 100, and some users are in both. "The communities are really important for both of these," says Saphra. "These are not tools you can use without being deeply involved with the community."

"These are not tools you can use without being deeply involved with the community."

Pimentel and Saphra are both from the United States, as are most of the coders they know who use voice recognition. The software is usually worse at interpreting speakers who have an accent other than standard American. According to an analysis (see go.nature.com/2ffx78z) by data scientist Rachael Tatman, at least some tools make more errors with women's voices. "If you don't have a mainstream American accent, or if you're a woman, then it's going to be a much more painful process," says Saphra. "But even then, it's not so bad."

There are other downsides. Pimentel has throat problems and has to take frequent breaks. "I drink so much frickin' water," he says. He's looking into voice training to learn to put less strain on his voice.

Pimentel misses being able to work in a quiet library, while Saphra misses being able to make more noise. "I used to listen to music or sing to myself while I coded. Or I would just curse. I can't do that anymore," she says.

But they appreciate the benefits, too. "I've often thought that if I woke up one day and my hands were miraculously perfectly fine, that I would continue to dictate code," Saphra says, only typing if something was tricky to dictate. "I think a lot of people could benefit from it." Voice command can even be relaxing, Saphra notes: she can put her feet up instead of hunching over a keyboard.

Probably the greatest extra benefit, though, is more intangible. "Now I really think carefully before I say anything or execute anything," says Pimentel, "and my code seems to have fewer bugs." Saphra adds: "I feel a mastery of my own tools in a way that I was never motivated to gain before this." ■

Anna Nowogrodzki *is a freelance writer based near Boston, Massachusetts.*

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

CAREERS

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

BLOG Personal stories and careers counsel **blogs.nature.com/naturejobs**

TWITTER For links to career resources and tips **go.nature.com/twitter**

ILLUSTRATIONS ADAPTED FROM VIKTORIA KURPAS/SHUTTERSTOCK



COLUMN

Harness the power of groups

PhD students can use writing meet-ups to beat isolation and depression, as well as to boost productivity and motivation, says **Karra Harrington**.

Feelings of depression, anxiety and isolation are so common during a PhD programme that some have dubbed the experience 'the PhD blues'. As a PhD student and practising psychologist, I wanted to try to reduce the impact of the blues on my fellow students and on me.

I decided to plan a regular meet-up with my student peers, in which we could write up our theses together. My hope was that it would establish deep social connections and help us to cope with some of the challenges of our PhD programmes.

I had attended Shut Up and Write! writing groups, which involve short 'sprints' of writing with breaks in between. Everyone works silently during the sprints and socializes during the breaks. These groups helped me to manage my productivity and motivation — but I had no sense of connection with my fellow writers. Often, participants would check e-mails or take a walk during breaks. And different people attended each session, which made it hard to get to know each other and to build connections. Any conversations were superficial and perfunctory.

For my group, I wanted to use the breaks to create supportive networks and to share ideas on how to overcome challenges. Ultimately, I wanted to create a community in which participants could learn from and support each other while also feeling productive and making progress on their theses.

I first needed to find participants. I reached out to students involved with the Cooperative Research Centre for Mental Health, an Australian research consortium based in Melbourne that aims to further mental-health research through collaboration. Students in the consortium work across research areas, institutions and geographical locations. Such diversity meant that relationships developed among students who would not usually interact regularly. It also meant that we could use the sessions to expand our networks and gain fresh perspectives on common challenges.

I launched the meet-up almost two years ago, and it has been a huge success, with ▶

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

a regular attendance of six to ten students every month. Our sessions are done in person and through videoconferencing, and include two to three hours of writing, as well as discussions on how to find mentors, structure thesis drafts or balance family life with completing a PhD. I facilitate each session, and group members raise topics according to their needs and interests. Members say that they feel accountable to the group, and that this motivates them and limits procrastination. They also check in with each other between sessions.



INTERACTION ZONE

Bringing together students from different institutions, and creating a space in which they could interact, was challenging. The videoconferencing helps members who can't get to campus — they feel engaged with their peers and less isolated. The meet-ups help me, too: I often have limited contact with other people during my working day. A regularly scheduled time to meet with others and discuss my science gives me a break from the isolation and is something to look forward to every month.

It took some time to build trust in the group so that everyone felt comfortable participating in the discussions. As facilitator, I keep discussions on track and relevant, encourage quieter group members to speak up and provide opportunities for those participating through videoconferencing to contribute. These efforts help to establish a sense of fairness and equality in the group.

The consistency of attendance by core group members is important because it helps participants to build relationships, thereby fostering a sense of safety and trust within each session. Group members say that they value having opportunities to connect with other PhD students and share their experiences, and that the group has helped them to maintain their motivation and sense of well-being.

Along the way, I've realized the importance of setting clear expectations using ground rules. Our ground rules are based on respect and confidentiality, and include speaking one at a time, listening to each other, not talking during writing sessions and maintaining confidentiality on all issues that we discuss in the group. All group members agree to stick to the rules, and the facilitator helps to enforce them.

As our group continued to meet and members started to open up about challenges that they faced, we found that the ground rules became even more valuable, because they helped to promote a sense of safety and encouraged useful exchanges between members. I found that it was also useful to make the group's purpose — to manage productivity and well-being — explicit from the start. On signing up to join, members know immediately what to expect. We use the Pomodoro Technique, which involves blocks of writing, breaks for discussion and goal setting, to manage our productivity. At the start of each session, we share our individual goals with each other; and during the discussion breaks, we check with one another on our progress towards our goals.

For some group members, this was a new way of working and it took some time to get used to. But the fact that we made it clear from the start why we had adopted this way of working helped new members to understand and to agree to try out these techniques. Members say they've found that the group helps them to set aside quality writing time, and that the structure of the sessions enables them to make progress on their thesis even when they are struggling with motivation.

The peer-mentoring aspect gives everyone a space in which to ask questions and to share what they know. A group member might, for example, seek advice about conference networking while also providing guidance on a data-analysis technique. By creating opportunities for members to ask for support and advice, the sessions help everyone to feel more hopeful and to identify proactive steps that they can take to overcome challenges. The opportunity to help others and to share knowledge provides everyone with a sense of empowerment and the ability to recognize their own strengths and expertise.

"To paraphrase Maya Angelou, my goal is not only to survive, but to thrive — with passion, compassion, humour and style."

Our discussions help to normalize the challenges of PhD studies and remind us to celebrate our successes. We wanted to share the benefits of the group with other PhD students, and so we've developed our model into a programme called Write Smarter: Feel Better. We have created guidelines for group sessions and training materials for group facilitators. These cover things such as how to build trust, and how to help the group reach agreement on ways in which members should interact with each other. Australia's University of Melbourne and Edith Cowan University are now testing the programme, with evaluations planned for completion by the end of this year.

At both universities, PhD students volunteer for the role of facilitator. We worked with the universities to develop strategies to support PhD students in this role; these included providing first-aid training in mental health and arranging for a university staff member to be a support contact. Importantly, although the universities offer support, the sessions remain led by PhD students and for PhD students.

Creating this meet-up group has been one of the most rewarding aspects of my PhD experience. I have learnt so much from organizing and facilitating the sessions, and now have a solid peer network. I have been able to gain insights on my research and career that I wouldn't have had if I had stuck to working on my own or only with lab peers. To paraphrase Maya Angelou, my goal is not only to survive, but to thrive — with passion, compassion, humour and style — and my meet-up is helping me to do exactly that. ■

Karra Harrington *is a PhD candidate at the Cooperative Research Centre for Mental Health in Melbourne, Australia.*

CORRECTION

The Careers Feature 'It takes more than a vow' (*Nature* **558**, 149–151; 2018) erroneously referred to the National Technical Institute for the Deaf as the National Technical School for the Deaf.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

DED-MEK

A vision of the past.

BY MATT THOMPSON

Ded-Mek slots the iridium strip into the boy's chest cavity and stands back to admire her handiwork. Her construction blinks at her, frowning. His skin grafts are almost perfect. Not too exact, of course. She always leaves imperfections, blemishes. Other than that you might be fooled; but Ded-Mek appreciates the glitches. It's as if Geon, her son, her ruin, is still growing up, still unfolding his petals of childhood to become the man he never will be.

The boy's cheek twitches. His eyes focus. His mouth opens.

They bring Ded-Mek their castoffs, their unloved and unwanted DNA strands, mock-orgs, memory chips. She scavenges in the ash-heaps that ring Nampo, she searches across New Korea, rooting through mountains of obsolete components and antique chromo-filters until she finds something she can incorporate into her latest version of Geon.

"He's not coming back, Ded-Mek," they say to her. "You have to let him go."

They say it with sympathy; but Ded-Mek pushes them out the door anyway, her face contorted into a fury so pure and profound none wishes to question her motives ever again. She's happy to be called by such a name. For her, the mech lives on forever.

As does her son.

Geon, her muse; Geon, her mystery. How well the citizens of this metropolis know his countenance. How they long never to see him again.

The boy staggers to his feet and lurches for freedom. Ded-Mek, her face shining, stands aside to let him pass. His hand brushes hers. A snap of static arcs between them. Already his stride is compromised, weak. An odour of cordite hangs around him, a stench of built-in obsolescence exuding from his pores like poison.

The sun sears this derelict suburb. Ded-Mek shields her eyes. The boy, as if following some primordial instinct, heads for the arterial highway a kilometre away. His feet drag. In the relentless glare he resembles a drifting shadow, cut loose from its half-life and cursed to wander Nampo for all eternity.

But eternity here is counted out in minutes and seconds; and Ded-Mek, following discreetly behind, prays he will last at least until the sun hits the zenith.



ILLUSTRATION BY JACEY

Her first Geon survived only minutes. Patched together from abandoned body parts and crushed genes, he shivered and died in the same growth-pod that birthed him. Undaunted, Ded-Mek immediately set to work on the first of his brothers, cloning a new Geon from the salvageable parts of the old.

That one lasted long enough to reel a few steps across the floor of her workshop. She cut him open while he was still warm — his heart, beating its last, had already shrivelled like an old, dried-out fruit. She spliced his DNA with some modified genotypes she had been gifted, and set to making a new, improved number three.

He is number thirty-seven. Or thirty-eight — depending on what you count. He has, in his limited wisdom, decided to shelter beneath a cluster of berry trees, away from the scorching heat of the sun and the howling noise of the highway. Ded-Mek stands a hundred metres away, willing him to get back to his feet. A crow pecks at the ground near to him. She knows it has smelt blood, sensed death.

But her creation will not be fodder for the beasts. She moves closer and hurls a stick at the interloper. Startled, it flaps leisurely away, cawing its displeasure.

The boy looks up. But Ded-Mek turns from him; she wishes for her children to make their own way in the world, to snap the apron strings and fly free.

The boy trembles. His gaze films over, and his head droops onto his chest once again.

She dreams of Geon every night. She recalls his laugh, his tears, the long months nursing him to some parody of health. The final, desperate lunge into the surgical unknown. The sight of his pale, bloodstained face in the mortuary.

Others have asked her to reconstruct their own scourges: a child, a partner, a beloved pet. Always she turns them down. Her skills, honed in the chaos of the mid-century simboom, are locked on one task only.

Those she refuses harangue her, mock her. Crazy, obsessed, a zealot; she has heard it all. But some day her Geon will return. Some day he will walk these streets again, his joyful shrieks echoing from the overpasses, his golden skin glowing in the light of the noonday sun.

When the boy is dead Ded-Mek hoists him over her shoulder and carries his near-weightless body back home. The children of the neighbourhood stand stiffly to attention as she passes, as they always do, their serious eyes watching her all the way.

She lays the cadaver out on her bench and slices open the chest cavity. The lungs are dead, useless, as is the heart. The spleen may be of worth; she extracts it with practised assurance and freezes it for his successor. When she is done she incinerates the body with no more grief than she would feel for a bundle of old rags. Her thoughts are turning to the future. One of her clients told her he may have access to the cell clusters the university's cloning labs routinely discard. She calculates the cost; it will be within her budget.

When the flesh is melted away, Ded-Mek cleans out the oven, remembering how Geon would help her around the workshop before his decline, in the years before she lost him.

But she can find him again. Filled with energy anew, she activates her database and begins the long scroll through the sequence-streams in search of a suitable match. ■

Matt Thompson *is a London-based musician and writer. His work appears or is forthcoming at* Aliterate, Black Static *and others. He can be found online at matt-thompson.com and on Twitter at @24wordLoop.*

➲ NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.